

Lessons from BSE for public confidence

The scare in Britain over infections across species by prions was badly handled. Whether the research community can do more to prevent future public crises of confidence needs to be examined in the light of the past influence of interested parties.

LAST week's spectacular collapse of public confidence in British beef (see page 273) was triggered in a scandalous fashion. The announcement by a British government minister that a new strain of the human prion disorder, Creutzfeldt–Jacob disease, had been identified and that there is circumstantial evidence of association with past consumption of beef affected by bovine spongiform encephalopathy, was fair enough. But for the evidence to be unavailable at the time of the announcement maximized the scope for public alarm and bans by other governments. The scandal is that the panic was foreseeable, as was the need to have in place the scientific advice on probable risks that in fact only followed several days later. The original evidence remains unpublished.

One reason for the delayed publication was said to be the need for independent review of the evidence. That is indeed necessary before public exposure and, whatever their expertise, the advisory committees involved are too close to the government to be seen as independent referees. But external review can be conducted within a day or two in exceptional circumstances.

Just as questionable was the attempt by the government to suggest that responsibility for advice to the public lay with its scientific advisers. But it is the government that must first assure itself of the quality of the advice it receives on scientific knowledge and its limits, and then to guide the public on the basis of that advice while also considering its wider implications.

Following the initial announcements, journalists seeking more information found that the scientists involved had suddenly vanished. It is well known that British agriculture ministry scientists are tightly constrained in what they can tell the outside world (however formally) even in the best of circumstances. Here, however, their invisibility, and that of others within the advisory system, was appropriate. It is a well justified principle that researchers should not comment on results until the latter are publicly accessible, especially when, as here, an unguarded statement can lead to extraordinary damage and potential legal liabilities.

A government (and indeed a country) impaled on a pitchfork of scientific uncertainty is thankfully an uncommon sight. But in the longer term other public scares can be envisaged in the growing use of agricultural biotechnology — and enhanced in their potential impact by the loss of government credibility in last week's events. This is where the public's perception of risk may bite once again. In generating public confidence, statements of quantified risk can be essential. In the current situation that is impossible without more epidemiological evidence, more research on species barriers, or a much better understanding of mechanisms.

Prion research will now be strengthened. But funding agencies need to ask whether their programmes are sufficiently aggressive in tackling other uncertainties that may fuel public disquiet in the future. The United Kingdom's technology foresight exercise last year included examination of agriculture and, separately, the food industry. Rightly, that process has fed into research funding policy. But in both cases, the process was dominated by scientists, industrialists and civil servants — a similar grouping to that accused, by its past role within government regulation of agriculture, of giving too little attention to health risks and too much to the needs of farmers and commerce. Has enough been heard from those more critically minded in the interests of public health?

In its regulation of agriculture, the government (and those in other countries where suppliers' interests weigh heavily) should in future lay the burden of proof more on suppliers than on consumers. In that context, the closeted activities of the UK's Ministry of Agriculture, Fisheries and Food — indeed, its continued existence in its present form — may, after critical examination, turn out to be a scandal as great as any other. □

Through the budget fog

A projected catastrophe for science in the United States undermines government credibility

PRESIDENT Clinton's 1997 budget is a cautious document, which implicitly acknowledges the brevity of its own shelf life: there are no grand schemes here, since the president knows that the Congress would simply ignore them. The budget does, however, offer a partial answer to a major science policy question which the White House has been reluctant to answer. If non-military research and development spending falls by a third under the Republican plan to balance the budget by 2002, as is widely believed, what happens to R&D under Clinton's plan to do the same?

The bad news is that it will not fare much better. The science, space and technology line in Clinton's budget — which includes the National Aeronautical and Space Administration, the National Science Foundation, and some energy department research — would fall precipitously from \$17.9 billion next year to \$14.6 billion in the year 2000, before recovering somewhat. That's a real cut of more than a quarter, and almost a billion dollars less than the Republican proposal.

For health research and training — the National Institutes of Health, broadly speaking — Clinton would raise spending from \$12.6 billion next year to \$12.9 billion in 2000: a 7% cut, when inflation is taken into account. This figure is higher than that envisaged in the Republican plan — but that plan is out of date, and Congress is, for the second year running, seeking more money for NIH than the administration has offered (see news, page 277).

The Clinton administration and the Republican Congress, while holding substantially different positions on such items as technology and environmental research, are basically united in their professed intent of hammering science spending in order to balance the budget. So it is wrong for Al Gore, the vice president, to say — as he did last month — that Republican plans for science spending will lead to Japan surpassing the US in non-military science spending “at the turn of the century” while the administration has “a very sensible plan” to prevent this from happening. It has nothing of the kind.

If the United States moves to balance its budget without tackling the areas where it actually spends three-quarters of the money — defence and mandatory health and social security programmes — then science spending will indeed be crushed, along with much else of value that the federal government does. The best that can be said for Clinton is that he is even less sincere about doing this than is the Congress. □



'Mad cow' scare threatens political link between food and agriculture

London. A decade of ministerial denials ended abruptly last week when the British government reluctantly acknowledged a possible link between consuming beef infected with bovine spongiform encephalopathy (BSE) and Creutzfeldt-Jakob disease (CJD), the fatal neurodegenerative disorder in humans.

The government continues to insist that British beef can be eaten with "confidence", and that the risk of contracting CJD remains "extremely unlikely". But the crisis in public confidence caused by last week's announcement has already prompted calls for changes in the structure of the Ministry of Agriculture, Fisheries and Food (MAFF), as well as the way that its research is linked to public policy decisions. Both follow suggestions that MAFF cannot represent both the beef industry and consumers.

The government immediately banned the use of cattle and sheep offal in the animal food chain (it is already banned in feed for animals intended for human use). It has even been contemplating the destruction of a substantial proportion of the national dairy herd on the grounds that it could be harbouring BSE. But 20 countries, including most of the European Union, have still suspended British beef imports.

The government announcement, an untypically swift response to new research findings (see right), has led to panic both among the beef-eating public and alarm in Britain's £2 billion (US\$3 billion) beef industry, which could be facing collapse. It has also led to many questions about the government's handling of the entire BSE affair since it began in the mid-1980s, including the way that scientific evidence and advice is handled both within and between different government departments.

In the face of opposition claims of jeopardizing the nation's health, government ministers continue to insist that they have acted responsibly. One of the ministers at the centre of the crisis, Douglas Hogg, secretary of state for agriculture, says his responsibility to the public will always take priority over his duty to the agriculture industry. Stephen Dorrell, secretary of state for Health, has said that decisions relating to BSE are based on the advice of the government's 'independent' 13-member Spongiform Encephalopathy Advisory Committee (SEAC).

SEAC is at the centre of the latest, as well as previous 'BSE-related' incidents. Its

Silent summer? Dairy herds may still be culled.

members are drawn from a number of research disciplines, and periodically advise the government on BSE policy after reviewing the latest research. It was advice from

SEAC that led to a ban of the use in cattle-feed of sheep and cattle offal in 1989, on evidence that cattle developed BSE after eating sheep offal infected with scrapie.

Recently, 10 new cases of CJD, of an average age below — and unlike — previous cases, were notified to SEAC. The committee's members told the government that, "in the absence of any credible alternative", the new infections could be attributed to all the victims having eaten beef infected with BSE before the imposition of the 1989 offal ban. (On Monday, John Pattison, professor of virology at University College and Middlesex School of Medicine in London, and chairman of the advisory committee, confirmed that two further suspected cases were under investigation).

In the past, the government has pointed to the absence of clear biological evidence to justify its refusal to acknowledge a link between BSE and CJD. The fact such evidence has still to be found has led the government's critics to point out that an ►

BSE: the questions that need answers

London. The British government's announcement last week that there may be a link between bovine spongiform encephalopathy (BSE) and Creutzfeldt-Jakob disease (CJD) is based on reports of 10 cases of a previously unrecognized variant of CJD.

The 10 cases all showed first symptoms within the last two years, the average age being between 26 and 27, with two in their late teens, and one aged 42. Two patients are still alive.

The new strain seems to be a clinicopathological variant of conventional CJD, with an unusual pathology closer to that of the human disease kuru than that of sporadic or iatrogenic CJD. Furthermore, the clinical signs are more similar to kuru, with more ataxia than dementia, a characteristic that is also true of iatrogenic rather than sporadic CJD.

It is not clear that this new variant of CJD is, in fact, BSE. But the fact that its pathology has never been seen before, either in the United Kingdom or elsewhere, as well as the unusual age grouping — CJD is rarely found in individuals under the age of 30 — have both led the UK Spongiform Encephalopathy Advisory Committee (SEAC) to its conclusion that a new risk

factor has appeared.

According to SEAC, the best candidate for this new risk factor is possible exposure to the BSE agent before a ban on offal was introduced in 1989. But further experiments will need to be done to assess whether this is in fact BSE.

This uncertainty is linked to the fact that there remains a great deal of uncertainty about CJD, BSE and prion diseases in general. The exact nature of the infectious agent is mysterious. Once believed to be a virus, based on its extreme resistance to ionizing and UV radiation — only solvent and extreme heat can reduce infectivity — it was proposed as early as 29 years ago that the agent contained no nucleic acid (see *Nature* **214**, 764; 1967).

The 'prion hypothesis', namely that the infectious agent is in fact an abnormal conformation of a naturally occurring protein that transforms the normally harmless prion protein into a protease-resistant dangerous form, is now close to being generally accepted.

The host prion protein, whose normal function remains unknown, is necessary for infection to occur; mice lacking the normal prion protein cannot be ►

Food/agriculture link threat

►(from previous page) earlier comprehensive offal ban in the human and animal food chain would not have created the present panic, nor imperilled the beef industry.

Erik Millstone, a senior lecturer and food policy specialist at the Science Policy Research Unit at the University of Sussex, is critical of SEAC. The government, he says, used SEAC's cautiously-worded advice to delay taking effective action. SEAC members, he adds, should have spoken out, instead of remaining silent. "Two months ago, Dorrell said that the likelihood of a link between BSE and CJD was 'inconceivable'. But I didn't hear any scientist say, 'no minister, that is not correct'." SEAC members ideally, he adds, should also include representatives of, "not just industry", but also consumer groups.

Millstone is one of a number of commentators calling for the break-up of MAFF on the grounds that a 'superministry' cannot protect the interests of both the agriculture industry and consumers when there is a direct conflict of interest between the two.

Under the existing arrangement, the agriculture industry is very much the senior partner within MAFF, says David Millar, a lecturer in media studies at the University of Stirling, who heads a research group into the

political dimensions of the BSE affair. The food safety directorate, adds Millstone, "is the poor relation". This has been confirmed by Edwina Currie, a former conservative health minister, who complained last week that public health plays a subordinate role.

Both Miller and Millstone claim that MAFF directly influences the nature of BSE research in favour of the agriculture industry. Millstone points, for example, to the lack of a single MAFF-funded research project aimed at developing a test to determine the presence or absence of the BSE pathogen both in live animals and in carcass meat. Miller claims that one researcher involved in identifying BSE was explicitly prevented by MAFF from describing BSE as 'a scrapie-like disease'.

MAFF, says Miller, also carries considerable weight within government over issues that affect the agriculture industry. He points out, for example, that the ministry over-ruled the Department of Health in 1988 during a scare over salmonella in eggs, when Edwina Currie, as health minister, acknowledged the claims of many scientists "that most of the egg production in this country is now infected with salmonella".

Currie's statement led to plummeting egg sales and open conflict between the Depart-

ment of Health — concerned at the implications for public health — and MAFF, which was worried about the damage to the poultry industry. MAFF gained the upper hand, says Miller and thus took over the coordination of the government's response.

"Very senior sources in the department of health have reported that MAFF put 'intense pressure' on the chief medical officer to say, 'forget about it, she was quite wrong, eggs are safe'", claims Miller.

Both the salmonella and BSE sagas have prompted calls for the creation of an independent food standards agency and the recreation of an independent food ministry which existed until a merger created MAFF in its current form in 1955.

Tom Blundell, chief executive of the Biotechnology and Biological Sciences Research Council, which spends £2.4 million (US\$3.75 million) each year on spongiform diseases, says the current BSE affair "certainly justifies another look at separating the regulatory role of MAFF in order to restore public confidence".

He adds that attention should focus on why public concern reached such high levels; in particular, the idea "shared by the public, but not by scientists" that "a scientific statement is law". Blundell, who is active in promoting the public understanding of science, says "a feeling for uncertainty" is essential towards understanding a subject such as the BSE/CJD relationship. **Ehsan Masood**

►infected, as no additional abnormal form can be produced. Recent experiments have shown that it is possible to propagate the abnormal, protease-resistant form of the prion protein in cell-free systems, and that even 'strain' properties can be faithfully reproduced in such *in vitro* assays.

But the story does not stop there. We know that there is a genetic predisposition for susceptibility to prion diseases, and a paper in this issue suggests that other cells may also be involved in the neurotoxic effects of the abnormal form (see page 345).

Furthermore, very little is also known about how, in cases of BSE, the prion particle moves from the digestive system to the brain. One suggestion is that it travels across the gut, via the lymphatic system, to the spleen, and then via the nerves innervating the spleen to the spinal cord and brain. A possible involvement of the immune system is also suggested by microglial involvement in neurotoxicity.

The main questions now being asked are whether BSE can infect humans who have eaten infected beef, and whether these new cases of CJD are due to BSE. One experiment almost certainly under way at present, which would confirm whether these cases are due to BSE,

would be to inject infectious material from the CJD patients into mice, in order to discover whether the clinico-pathology resembles that already well described for BSE in mice.

The British government has pledged more money towards prion research. But where should that money go? Research

Trouble spots: brain section showing white lesions of Creutzfeldt-Jakob disease.

into the precise nature of the infectious agent must be continued; if scrapie protein produced in the cell-free system were shown to infect mice, the 'protein only' prion hypothesis would be almost certain. Other research priorities are the need for more work into the precise route of infection, from exposure to disease onset, and into the genetic

factors influencing susceptibility.

More practical studies could include the development of early diagnostic tests for both humans and cattle. Therapeutic approaches are also an important goal; while antifungal drugs and polyanions at high doses inhibit transmission in mice, the side-effects are undesirable, and attempts to treat CJD patients with such drugs have not been successful.

Better methods to detect infectious prion particles need to be developed. The present detection systems of Western blotting or bioassay, based on injection into mice, are not very sensitive; the 'species barrier' between cattle and mice requires the equivalent of 10,000 times more of the BSE agent to be injected into a mouse than into a cow to produce the disease.

Transgenic mice expressing the bovine prion protein might provide a more sensitive assay for BSE in potentially infected tissues. Furthermore, biochemical or other direct detection methods would be particularly welcome.

More needs to be done to assess whether these 10 cases are indeed due to BSE, and if so what the potential number of future cases might be. But SEAC remains convinced that measures taken over the past 7 years mean beef is safer now than ever. **Kimberly Carr**

Ralph Eagle/SPL

Biosphere 2 begins fight for credibility

Oracle, Arizona. Columbia University is planning to use a modest programme of carefully-controlled experiments in biogeochemistry and plant physiology to rebuild the scientific credibility of the Biosphere 2 complex, for which the New York-based university assumed full management responsibility at the beginning of the year.

The programme will mark the end of efforts to use Biosphere 2 to replicate an ecosystem parallel to that of planet Earth. Instead, scientists using the facility will pursue a more prosaic goal, namely the conduct of properly controlled experiments on a par with others already taking place in ordinary greenhouses around the world.

Plans to prepare such a research programme were endorsed by an informal panel of two dozen external scientists brought together by Wally Broecker, professor of geology at Columbia's Lamont-Doherty Earth Observatory (LDEO), at meetings held at Biosphere 2 two weeks ago.

But many panel members continued to express scepticism about how useful the giant, fully-enclosed greenhouse is likely to be for scientific research. As an ecosystem, it bears no obvious relation to the outside world. Its soil is unusual, its lighting — blocked by the over-engineered structure — is poor, and there is no wildlife except for ants and cockroaches.

Furthermore the ability to 'close' the system is incomplete, and of limited scientific value. As a greenhouse, the Biosphere is dark and expensive to run. Indeed its very uniqueness makes replication difficult.

Nevertheless, according to Bruno Marino, the facility's chief scientist, Biosphere 2 presents "a fantastic opportunity to look at the pattern of response of different plants to shared environmental conditions". Most panel members agreed that, given sufficient determination and imagination, a way can be found to do good research there.

"The Biosphere was not built to do science, and its conversion is going to be tricky," concedes Broecker, the man behind a deal which Columbia has struck with Ed Bass, the Texan billionaire, to run the complex at his expense for the next five years (see *Nature* 378, 223; 1995). Bass had previously spent up to \$150 million building the complex, which was used in two notoriously failed attempts to enable humans ('Biospherians') to survive in complete isolation from planet Earth ('Biosphere 1').

The panel considered, in essence, two opinions of how the complex should develop. Broecker wants to use it for simple, quantifiable experiments, such as comparing the progress of plants at two different carbon dioxide levels. Biosphere ecologist Tony Burgess, however, advocates a qualitative

programme, closer to Biosphere's original concept, to study biodiversity in a self-contained ecosystem.

The ecologists think that the complex can serve as a model for the human-controlled ecosystems of the future. Biosphere's problem of managing a small, artificially constrained ecosystem "is exactly the problem that is facing every director of every wildlife reserve in the world," says Burgess.

IMAGE UNAMIGABLE FOR A CORN AGRICULTURE FOR REASONS REASONS

Biosphere/Gill Kenny

But Broecker says that he cannot see how the ecologists would use the Biosphere to yield novel scientific information. "You're not really saying how you're going to use that ecosystem to solve any important problem," he told the ecologists. Broecker has a major problem of his own in mind, namely the increase in carbon dioxide in the atmosphere, and the study of its impact on plants. "Our thrust is to utilize Biosphere 2 in that quest," he says.

Work is already under way on this topic. Guanghui Lin, a researcher at Biosphere 2, has already used the carbon dioxide build-up there — caused by the compost-like soil — to monitor the responses of different types of plant.

The panel endorsed Broecker's approach. But the layout of Biosphere 2 will still allow some ecological work to be done. The tee-shaped greenhouse consists of two

chambers: the "wilderness" chamber and the "agriculture" chamber where the Biospherians used to try to grow their food.

Under the \$150,000 plan endorsed by last week's panel, the agriculture chamber will be split into three sections this summer. The separation will be by polythene sheeting: solid glass walls would be expensive, light-reducing, and vulnerable to pressure gradients. As a result, scientists such as Lin will be

able to run proper control experiments, improving their chances of having their results published.

The panel also suggested that the agriculture area — much of which is now full of wheat — should grow some kind of tropical agroforest system, such as coffee and inga, which would take advantage of the greenhouse's height.

The wilderness area will continue to attract the interest of tourists. But its future as a research tool is as murky as its artificial ocean, algae-filled and bereft of other life. "I was disappointed that there were no really crisp ideas" for biodiversity research in the area, says Broecker.

Columbia will review its participation after three years, according to John Mutter, interim director of LDEO. "We fully expect to be here beyond five years," he says, adding that the objective is to break even in that time.

Tourist income already covers \$2 million of the complex's \$6 million annual running costs. Columbia expects to make up the rest chiefly through educational programmes, starting this year with a summer camp and a fall semester for paying undergraduates.

Research will underpin this intellectually, but not financially: Biosphere 2 will not attract a huge amount of outside money, says Mutter.

Colin Macilwain

Mandela backs role of academy

Cape Town. South African President Nelson Mandela has challenged the country's newly-formed Academy of Science to inspire the country's youth to seek careers in science, engineering and technology.

Addressing the inaugural dinner of the academy in Pretoria last Friday, Mandela described as "immense" South Africa's need for rapid expansion of its scientific and technological skills. "None of us need to be persuaded of the utility of science to national growth and prosperity," declared Mandela. He emphasized that the academy had the potential to play a critical role in the development of both science and the nation as a whole.

In particular, Mandela invited the academy to comment on the draft white paper on science and technology, which is

due to be released during the course of the current parliamentary session. He also asked the academy to back closer cooperation between South African scientists and other African scientists, in order to share the former's skills and technology with the rest of the continent.

The new academy has elected Khotso Mokhele, a microbiologist who is the president-elect of the country's main funding agency for research, the Foundation for Research Development, as its first president. Jennifer Thomson, of the University of Cape Town, and Njabulo Ndebele, vice-chancellor of the University of the North, have been elected as vice-presidents, and William Makgoba, of the University of the Witwatersrand as general secretary.

Michael Cherry

CSIRO is streamlined under new boss

Sydney. The new chief executive of Australia's giant Commonwealth Scientific and Industrial Research Organisation (CSIRO) has announced a number of key changes to the organization, including the removal of a complete layer of management.

After little more than a month as head of Australia's largest employer of scientists, Malcolm McIntosh has said that he intends to reshuffle CSIRO's basic operating units — its 36 divisions, employing 7,000 staff. High among those changes will be a reshaping of the Division of Animal Production, which McIntosh says would have been declared "insolvent" if it had been in the private sector.

In his first major announcement since taking control of the CSIRO in February, McIntosh said last week that its six institutes — an intermediate level of management between the divisions which carry out research, and the organization's headquarters — would be abolished on 1 July.

Four of the six current institute chiefs will become deputy chief executives and sit on an executive committee with McIntosh. This will meet several times a month, in contrast to the existing executive committee, made up of institute chiefs, which meets just once a month. The two other acting institute chiefs will return to other duties and await vacancies on the executive.

In addition to sitting on the executive committee, the four deputies will head loose groupings of divisions, called alliances, which will promote cross-divisional research. Although similar to the existing institutes, the alliances will not have any administrative

structure or secretariat. All staff now employed by the institutes will move to CSIRO headquarters, or to the divisions.

The abolition of the institutes has long been expected. Originally created about a decade ago to improve the coordination of the activities of CSIRO's divisions, they have since been widely condemned by staff, as well as a 1994 government inquiry, for simply adding to the organization's bureaucracy and paper work.

An internal management report to the CSIRO board in 1995 also recommended abolishing the institutes, among other changes, although it did not accept that the institutes had caused any difficulties. But the then acting chief executive, Roy Green, and the board declined to take any major action until the new chief executive had arrived.

The appointment of McIntosh, a physi-

cist and previously chief of defence procurement in the United Kingdom, was announced in July last year but he was only able to start his new job in February. He now faces the formidable task of rejuvenating an organization that is not only believed to be awash with paper work, but has recently seen a number of disputes between divisional research staff — including the divisional chiefs — and the CSIRO board.

The changes announced last week have been broadly welcomed by staff, although some have expressed reservations over the fact that the executive committee will include the same senior executives as previously, and concern that the alliances could still generate a considerable amount of bureaucratic paperwork.

One division that may face significant changes, including staff redundancies — there have been reports that between 60 and 120 staff are thought to be under threat — is the Division of Animal Production. The division's A\$13 million (US\$9.6 million) a year budget relies heavily on funds from the wheat and sheep farming sectors. But with the collapse in wool prices two years ago, the wool sector has fallen on hard times, and such funds have dried up.

MacIntosh says that he hopes all changes immediately required will be finalized by the end of the financial year, but adds that they will not be the last changes he will make. He adds that the divisional structure should evolve as circumstances change, and that mergers, demergers and even staff retrenchments could become a permanent feature.

Mark Lawson



US tops international league in action on blood products

Montreal. A survey by members of the law faculty at the University of Toronto of how seven industrialized countries — excluding Japan — responded to the HIV-tainted blood crisis in the mid-1980s rates the United States best and Switzerland worst in the promptness with which they implemented the two most important risk reducing measures into their distribution.

They conclude that the performance of all seven countries "was far from exemplary, and exhibited serious institutional shortcomings leading to many readily avoidable deaths". This was due partly to social and cultural attitudes, including the view that AIDS was a disease confined to homosexuals, intravenous drug users and sexually promiscuous members of minority ethnic groups.

Outside the United States, AIDS was also viewed as primarily an "American disease", says that study. Only when all users of the blood system were "widely perceived to be at serious risk" did the institutions begin to respond effectively to the crisis.

In every country studied, institutions centrally involved persisted in denying risks of HIV contamination of the blood system well beyond the time when their existence had been scientifically proved. In many cases, "the vigour of these denials increased and persisted as a direct function of the extent to which the blood system was dominated by non-profit institutions," says the report.

The United States introduced heat treatment of blood products in October 1984 and the ELISA HIV screening test in May, 1985. In contrast, Switzerland did not use heat treatment until May, 1986, and the ELISA test until November 1985. Canada shared the worst-place rating with Switzerland in terms of the introduction date of ELISA.

The preliminary results of the study, carried out by Michael Trebilcock, Robert Howse and Ron Daniels of the law faculty's Centre for the Study of State and Market, were presented to a meeting at the university earlier this month. The full results are due to be published in the *Virginia Law*

Review next autumn.

The survey's purpose was to attempt a comparative analysis of the performance of the sharply differing blood systems of the countries involved.

"Our aim is not to make an overall judgement about these systems, either against efficiency or other criteria, but to understand their capacity to respond to a crisis that threatened the integrity of the blood supply," say the authors of the study.

The authors say that British and French governments, as owners and operators of elements of the blood system, seemed much more protective of them than governments that had a less direct stake in their systems.

In Canada, where federal regulatory responsibility was well established, the financing arrangements entailed complex and interprovincial negotiations, and considerations of regional development, job creation and industrial policy, that eventually accomplished nothing and wasted millions of dollars, says the report. **David Spurgeon**

Energy document lists laboratory roles, not goals

Washington. The operations of the Department of Energy (DoE)'s laboratory network have been spelt out in unprecedented detail in a draft "strategic mission plan" that was released by the department in Washington last week.

But Charles Curtis, deputy energy secretary, warned that the document fell considerably short of the department's goal of creating a mission plan for the laboratories, which, he said, would take time. "To make this into a true, living strategic plan for the laboratories is going to take the next year or so," he said. The draft plan, he said, "describes how things are. It does not describe where we should go."

Curtis also admitted that the detailed information in the plan could be construed as "an invitation to make mischief", as it contains much explicit information on the exact role of the laboratories. About 150 different programmes in both nuclear weapons and non-weapons areas across the department are described, along with details of how money for each one is distributed between the laboratories.

To take one example at random, the plan explains that the DoE spends \$2 million on research into carbon dioxide in the atmosphere, with 90 per cent of it to the Oak Ridge Laboratory in Tennessee. Pressure groups opposed to such research will be now be able to focus their protests on the laboratory, safe in the knowledge that laboratories outside are not directly involved.

The plan — which was demanded last year by a commission, chaired by the industrialist Robert Galvin, which investigated the future of the laboratories, and drawn up by the new Laboratory Operations Board — contains a new mission statement for the laboratories.

It also seeks to categorize each laboratory's role in each mission, identifying "principal laboratories" that do more than 20 per cent of a programme, "major contributing laboratories" that do more than 10 per cent, and "specialized participating laboratories" that do less than that. But no effort is made to change or set directions for these roles.

Some members of the Secretary of Energy's Advisory Board, which received the plan last week, complained that it lacked information on how the laboratories should make better use of industry and the universities for research — a charge which Curtis accepted. "Other things being equal," he said, the department should "favour the universities" as a place to do research, because of their teaching role. "We are not just in the business of defending the federal laboratories."

Colin Macilwain

NIH clinical research plans under fire from Republicans

Washington. Republicans in the US Congress have attacked President Bill Clinton's 1997 budget proposals for the National Institutes of Health (NIH), saying that the administration's plans outlined in the budget to build a new clinical research facility would be at the expense of grant money for individual scientists.

John Porter (Republican, Illinois), the chairman of the House of Representatives appropriations subcommittee that oversees the NIH budget, last week called the president's proposal "phony" and "irresponsible", and said that the level of new research money it requests is "clearly not acceptable." Porter added:

A spokeswoman for Porter's counterpart in the Senate, Arlen Specter (Republican, Pennsylvania), said that the Senator had similar concerns about the impact of this decision on the extramural research programme. "[Specter] is supportive of the Clinical Research Center, but he's concerned because so much of the [new] money goes there," said Margaret Camp. Specter chairs the Senate Appropriations Committee subcommittee which oversees the NIH.

Once the \$310 million that the president requested for the new 250-bed clinical facil-

what's the problem with a Clinical Research Center?"

John Gibbons, the director of The White House's Office of Science and Technology Policy emphasized at a briefing last week that there was a need to "accelerate" building the CRC to replace a decrepit 1953 facility. Harold Varmus, the director of NIH, argued separately that the president's budget still provides a shot-in-the-arm for individual investigators. "We have enough of an increase for us to pay more grants in 1997. The important thing is keeping the number of new awards up," he said.

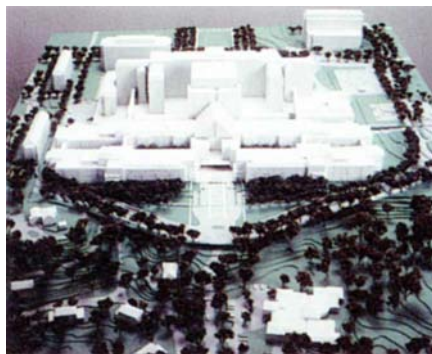
In the budget request for the fiscal year beginning 1 October, sent to Congress on 19 March, President Clinton asked for \$467 million in new funding for the NIH, boosting the biomedical agency's 1996 budget by 3.9 per cent, to \$12.41 billion. Sixty per cent of the new money would be used to build the CRC, an 850,000-square-foot addition to the NIH's existing clinical facility, parts of which would be converted to other purposes. Completion is planned for early next century.

The administration proposes that most (86 per cent) of the rest of the new money be used to support non-NIH scientists, providing with a 2.6 per cent increase overall in biomedical grant funding for non-NIH scientists. This would include 207 new research project grants. Intramural scientists, meanwhile, would lose 0.2 per cent of their 1996 funding, while research and development contracts would fall by 0.9 per cent.

But even some of the external scientists warn that laying out \$310 million for the CRC in a single fiscal year may cause an uneven flow of funds for multi-year grants downstream. "Instead of having a smooth flow you have something of a roller coaster," says Ralph Bradshaw, the president of the Federation of American Scientists for Experimental Biology, which last month recommended a 6.5 per cent increase in NIH funding over 1996. Spreading the \$310 million for the CRC over three years would probably be "more to our liking."

Asked whether he supported the size of the \$310 million request for the CRC, and its disbursement in a single fiscal year, Porter said that the main purpose of hearings was "to get at the authenticity and credibility of the figures." The subcommittee on labour, health and human services, and education of the House Appropriations Committee has scheduled a hearing on the CRC for 23 April. The Senate subcommittee plans to begin hearings on the NIH budget on 16 May.

The president's budget proposal also calls for the Office of AIDS Research (OAR) to maintain control over the NIH's \$1.4 billion AIDS research budget. **Meredith Wadman**



Clinical model: centre as envisaged by Zimmer Gunsul Frasca Partnership.

ity is deducted from next year's budget request, the rest of NIH would, under the president's plan, receive an increase of only 1.6 per cent over 1996. Porter said that as an alternative, he would seek a 6.5 per cent increase in research funding, in addition to the money earmarked for the new Clinical Research Center (CRC).

But Clinton administration officials robustly defend the request. "The president has been very aggressive in trying to bolster NIH research, and the 1997 budget reflects that," says Lawrence Haas, a spokesman for the Office of Management and Budget, pointing out that Clinton has budgeted hundreds of millions for the clinical centre "in an era of very scarce resources". He adds: "If the other side is concerned with research,

Berlin universities mired in budget crisis

Munich. An emergency budget bill was due to be debated in Berlin's parliament this week proposing a cut of nearly DM200 million (US\$137 million) in the overall budget of the city's three universities. It says that this cut should be achieved by closing or merging a number of specific faculties, and by reducing student numbers by 15 per cent.

The Berlin legislature was also scheduled to debate amendments to the city's universities law which would give its government the power to enforce faculty closures against the wishes of the universities. But the universities have reacted angrily both to the threat to their constitutional autonomy — under which they are in principle free to make their own academic decisions — and to the extent of the cuts.

The universities point out that, when added to previous cuts imposed since reunification in 1990, the total cuts are equivalent to the costs of running a whole university and medical school. They also emphasize that any enforced reduction in student numbers would conflict with another constitutional right: that of students with a school leaving certificate to a university education.

But, far from being united by their shared adversity, the universities — the Humboldt in the east and the Free and Technical universities in the west — have been thrown even deeper into the conflict that has been raging in Berlin ever since the full economic costs of reunification first became apparent.

Hostilities between the three universities began with the passing of the universities law in 1993. This predicted a lowering of the total universities budget by DM130 million within ten years, and a reduction in the number of students from 115,000 to 100,000. At that time the two west German universities were asked to bear the brunt of the savings, while considerable investment was made in the Humboldt University to raise it to western standards. The Humboldt, however, was required to halve its staff.

Last year, as the city's financial crisis deepened, the target for budget reductions was increased to DM288 million. The government wanted the cuts to be achieved

through agreed mergers of departments and faculties whose courses were duplicated.

But the universities have been very slow to reach agreements. Negotiations between them became increasingly hostile, and positions stubbornly entrenched. As Berlin plunges towards bankruptcy, the proposed cuts bring to DM484 million the total saving on annual budget expected to be achieved by 2003, and the government does not want to wait much longer for universities to decide on how they should be implemented.

The government's new proposal to parliament would allow universities to make suggestions for implementing cuts. But, if they do not do so within a defined time, it would enforce its own suggestions. The universities claim that the faculties targeted by the government for closure have been chosen randomly, but both the Technical University and the Humboldt University are refusing on principle to suggest alternatives.

Gunther Kaindl, vice-president of the science and research faculty of the Free University, which must cut back around DM40 million per year, says this attitude is not helpful. The Free University senate plans to offer an alternative to the plan to close the informatics department, with a saving of DM4 million, through restructuring of mathematics and materials sciences disciplines.

But such cooperation has its own limits: the university has no alternative suggestion to the closure of its school of dentistry which would save DM15 million. Kaindl points instead to the expensive new chemistry department, being built for the Humboldt University on a green-field site, when, he says, both west Berlin universities already have large and modern laboratories. "This all makes discussions between the universities very difficult," he says.

According to Bert Flemming of the social democrats — one of Berlin's coalition partners — a possible solution to the problem of restructuring could be to set up a committee made up of representatives of all three universities and the ministry for science and research, to make common decisions.

Up to now, he says, the universities have been opposed to such a committee, as it would allow outsiders to become involved in decisions that should be made internally. But, faced with the larger threat to auton-



Taking a stand: Humboldt University is not volunteering cuts to Berlin's government.

omy posed by the amendment to the universities act, he says, they may be more willing, and has been lobbying hard among his parliamentary colleagues for this solution.

Some moves towards this system of decision making are already introduced in the emergency budget bill. This would require universities, in discussion with politicians and community representatives, to agree on how to share between their science faculties an additional cut of DM14 million, not earmarked by the government for particular departments.

A similar committee is expected to distribute a 20 per cent cut on Berlin's medical faculties, which have suffered particularly from the post-reunification restructuring (see *Nature* 369, 431; 1994). Peter Gaehtgens, the Free University's vice-president for medicine, reluctantly accepts this principle, at least as a short-term measure. "It is quite apparent that the universities cannot come to a solution which is compatible with the economic situation", he admits.

But universities are unlikely to want the system to be extended beyond particularly difficult — or expensive — disciplines, he says. "That would be too intrusive".

The issue of student number reduction is equally problematic. Earlier this month, a Berlin court ruled in favour of 14 students who had been refused admittance to Berlin's medical schools because numbers had been reduced in previous cuts, agreeing that this conflicted with their constitutional right to study. The science and research ministry is now considering taking the matter to the constitutional court to clarify whether economic difficulties can be allowed to override this right.

Alison Abbott

Japan agrees to pay HIV-blood victims

Tokyo. After a seven-year court battle, hundreds of Japanese last week finally reached a compensation settlement with the government and pharmaceutical companies over their infection in the mid-1980s by blood coagulants that had not been heat-treated to kill viruses.

The plaintiffs, which consist of two groups in Tokyo and Osaka, decided to accept the out-of-court settlement after the Ministry of Health and Welfare and

the five blood product manufacturers they were suing formally apologized and admitted responsibility for the disaster.

Days before the settlement was reached, the Tokyo District Prosecutors Office announced that it had set up a team of six prosecutors to investigate possible criminal charges. Prominent among those who are to be investigated is Takeshi Abe, who headed an AIDS study group which was set up in the ministry in June 1983.

D.S.

HUGO approves ethics code for genomics

Heidelberg. Genetics researchers should not offer "undue inducement" to individuals, families and population groups taking part in gene mapping and other related experiments, according to a code of ethics for genomics research approved last week by the Human Genome Organization (HUGO).

But the code does endorse various types of agreement that might be reached in exchange for such participation, ranging from the provision of health care or "information infrastructures" to "the possible use of a percentage of any royalties for humanitarian purposes".

At the same time, it underlines the importance of obtaining informed consent for such participation "free from coercion by scientific, medical or other authorities". Such consent, it adds, can be "individual, familial or at the level of communities and populations" — a potentially controversial statement, as it does not require that consent be expressed by an individual participant.

The code was endorsed last week by the executive council of HUGO, the umbrella organization which loosely oversees both the Human Genome Project (HGP) and the Human Genome Diversity Project (HGDP). It was immediately distributed to members of the organization attending its first full international meeting held in Heidelberg in Germany.

Grant Sutherland, director of the department of molecular genetics at the Women's Hospital in Adelaide, in Australia, and president of HUGO, says that the organization felt it was important to make a statement about ethics that tried to be culturally sensitive, but without giving in to the dictates of any one particular culture. "The code is not law, and it cannot be imposed on anyone," says Sutherland. "But we believe it is a useful starting point for anyone in this area."

The working group which drew up the statement says that its recommendations are based on four principles: recognition that the human genome is part of the common heritage of mankind; adherence to international norms of human rights; respect for the "values, traditions, culture and integrity" of participants; and "acceptance and upholding of human dignity and freedom".

The sections of the statement likely to attract the most attention, following recent controversy over the implementation of the HGDP, are those dealing with the need to ensure that the informed consent has been obtained from participants, and for the provision of appropriate compensation (see *Nature* 377, 373; 1995).

But the first principle listed in the statement concerns the way that research should be carried out, specifying that "scientific competence is the essential prerequisite for ethical research".

"It can no longer be assumed in this field

that there is sufficient quality control for public safety always to be taken into account in genetic testing programmes," says Bartha Knoppers, professor of law at the University of Montreal in Canada, and chair of the working group which drew up the statement.

Knoppers also defends the decision not to specify that consent must always be at the level of the individual — standard practice in Western countries — on the grounds that in some societies, such decision are normally taken "at the level of the whole community."

Furthermore, although acknowledging that the working group which drew up the ethics statement has refrained from criticizing any specific practice, Knoppers argues that more specific judgement needs to be based on a clear agreement on underlying

principles. "Rather than working backwards, we have a chance to prospectively guide countries so there is no 'shopping' and 'tourism' by geneticists between different locations," she says, a reference to fears that a patchwork of different guidelines would lay populations open to exploitation.

Eric Lander, director of the Whitehead Institute at the Massachusetts Institute of Technology, suggested that, as well as being agreed to by publicly-funded researchers, the principles should be publicly endorsed by biotechnology companies engaged in searching for potentially useful genes among human population groups. "I would urge companies to say whether they sign on [to these principles] or not," says Lander.

David Dickson

NIH seeks rapid sequence release

Heidelberg. Intense negotiations are taking place between the US National Institutes of Health (NIH) and six anticipated recipients of grants for large-scale genome sequencing over the speed with which these centres should place the data obtained from their efforts in the public domain.

The National Center for Human Genome Research (NCHGR) is due to announce the allocation of \$15 million earmarked by NIH for advanced sequencing projects within the next ten days. Francis Collins, the director of the centre, is seeking a commitment from grant recipients that their results will be rapidly disseminated.

In particular, Collins wants the centres to agree to adopt release procedures as close as possible to the terms of an informal agreement reached by representatives of sequencing centres from the United States and Europe in Bermuda in February.

These suggest that a desirable goal would be for assembled sequences greater than 1 kilobases to be released daily — a practice already adopted by the UK Sanger Centre, funded jointly by the Wellcome Trust and the Medical Research Council, as well as Washington University in St Louis.

John Sulston, director of the Sanger Centre, defends this policy on the grounds of the need "to avoid unnecessary competition and duplication in the human genome community", suggesting that the most important result of rapid release of data is that it promotes openness between researchers.

Others sequencers, however, are less convinced by such arguments. While

accepting the principle endorsed at the Bermuda meeting — that sequences developed with public funds should be placed in the domain with the minimum delay — there remains strong disagreement over how short this delay must be.

Some argue that the published sequences should be carefully checked for accuracy before release; others that they should be scrutinized for potentially patentable information. In both cases, say some critics, the type of immediate release favoured by Sulston could have unforeseen, and potentially damaging, implications.

Those who have reservations about Sulston's approach include Craig Venter, the director of the Institute for Genomic Research (TIGR) in Rockville, Maryland, which has high expectations of being among the recipients of NCHGR funds for large-scale sequencing.

Venter says that he is prepared to abide by NIH policy on the question of the speed with which the sequences are released to the public. He also agrees that it is important to get sequencing data out to scientists as soon as possible. "But I feel that it should be accurate and clean," says Venter.

Collins, meanwhile, says that the NIH is keen that the agreements reached with sequencing centres reflect the philosophy of the informal agreement reached in Bermuda without violating the terms of the Bayh-Dole Act.

Sulston remains committed to the principle that, in the absence of further information about biological function, human genomic sequence information should be regarded as "pre-competitive information".

D.D.

France accused of 'reneging' on promises

Paris. The French labour union representing scientists, — le Syndicat National des Chercheurs Scientifique (SNCS), last week called for the resignation of Guy Aubert, the director general of Centre National de la Recherche Scientifique (CNRS), over his handling of a funding crisis at the agency.

The SNCS call came during a week of protests over a long-running financial crisis at CNRS and other public research agencies. Around 1,500 researchers staged a demonstration outside CNRS's headquarters in Paris.

CNRS researchers are angry, in particular, at what they claim to be the brutal way in which the agency's management confiscated the balance on laboratory accounts at the beginning of the year — by computer transfers carried out during the night — without warning researchers beforehand. This operation left some laboratories without the funds they had set aside for long-term programmes (see *Nature* 380, 189; 1996).

CNRS officials have since promised to reimburse some funds that, they argue, were taken "by mistake". They have also said that funds will be advanced to laboratories that have found themselves with no money for 1996. But the agency has not said when it will make these funds available.

Adding to researchers' concerns, CNRS

officials last week told a meeting of the agency's board last week that they could not guarantee that the agency would be spared from further cuts that may have to be made as part of a broader government plan to freeze FF20 billion (US\$4 million) of public spending this year.

Many scientists are claiming that the government has reneged on its promise to repay money owed to CNRS as a result of earlier budget decisions. Edouard-Henri Audier, for example, a chemist at the Ecole polytechnique, and a member of a group of laboratory directors that has just been set up to coordinate action against the cuts, argues that while the government increased CNRS's budget for running costs by FF 300 million last year, a similar sum was deducted at the same time from the salaries budget.

Researchers say they acknowledge that CNRS officials are faced with a difficult task in balancing the agency's books. But they criticize the agency for apparently supporting what they claim to be the government's failure to honour its commitments to research. They are also concerned about a freeze in recruitment. This year, public research organizations will create only eight new posts, with others being recruited in line with retirement rates — currently around 2 per cent. INSERM, for example, will only

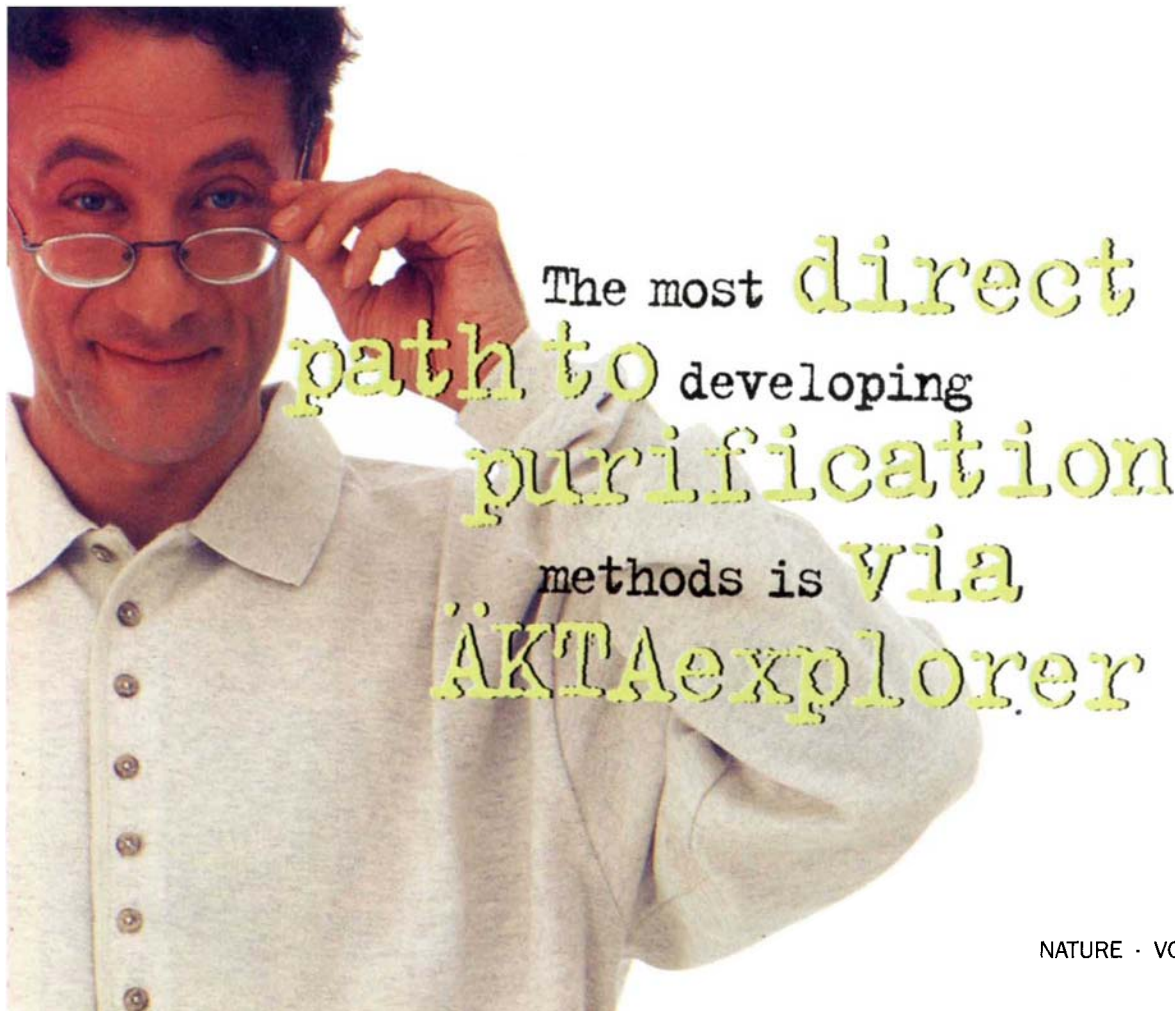
IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Back on the streets: CNRS researchers hold a protest meeting over government cuts.

recruit 36 researchers this year, compared with 76 in 1993.

The freeze in recruitment could not come at a worse moment. Over half the staff of the research agencies will retire before 2005, and recruitment needs to be reinforced now if a sensible age pyramid is to be maintained. "It is not in our interest to wait until then to recruit massively" says Philippe Lazar, director general of INSERM, adding that this would inevitably also lead to a lowering of the entry standards required.

Catherine Tastemain



The most **direct**
path to developing
purification
methods is **via**
AKTAexplorer

Democrats lash out at majority members on House panel

Washington. Democrat members of the Science Committee of the US House of Representatives have launched a withering attack on the Republican majority on the committee, alleging that "good science is being smothered in its crib because the majority fear what that science may tell us when it grows up".

The usually staid *Views and Estimates* document on the budget — just released by the staff of ranking minority member George Brown (Democrat, California) — says that the majority "failed to do [its] work in 1995", holding only 46 hearings, against 113 in the first year of the last, Democrat Congress. "This Committee is crippled in its ability to provide a responsible analysis of the direction and funding of our programmes," says the document. □

Global warming hits insurance

Munich. Munich Re, one of the world's largest reinsurance companies, last week announced that the economic cost of natural disasters rose last year to US\$180 billion, almost three times that of the previous record year, 1994. Although around US\$100 billion of the 1995 costs were due to the Kobe earthquake, the company is asking for political action to introduce sustainable measures to control global warming, which it blames for the increase in disasters. □

Deal backs contract researchers

London. Britain's main research funding agencies have reached an agreement with the Committee of Vice-Chancellors and Principals (CVCP) providing a 'framework of principles' within which individual institutes will be expected to build their own policies aimed at improving the career prospects of research staff employed on short-

term contracts in UK universities.

These principles, the acceptance of which will in future be requisite for the receipt of grants from the research council, emphasized that all institutions have a responsibility to promote the careers of contract researchers, and must accept the importance of regular review and career guidance for such individuals. The agreement also includes measures to support paid maternity and sick-leave arrangements. "This is a major step in promoting equal opportunities in university employment," Diane Warwick, the chief executive of the CVCP, said last week. □

Wellcome funds TB sequencing

London. Britain's Wellcome Trust has announced that it is to support a project to sequence the complete genome of *Mycobacterium tuberculosis*, the organism that is responsible for more deaths world-wide than any other single infectious disease.

The project, announced on last week's TB day organized by the World Health Organization, will be based at the Wellcome Trust Genome Campus at Hinxton, near Cambridge. In keeping with the policy of the Sanger Centre, which is based on the campus, the first data from the project have already been available via the world-wide web at <http://www.sanger.ac.uk>. □

Indian launch boosts independence

New Delhi. The prospects of India entering the international launch business brightened last week when its polar satellite launch vehicle (PSLV) placed a 930-kg remote sensing satellite in a polar orbit after blasting off from an island off the west coast.

It was PSLV's third flight, and its second successful one. Officials say that PSLV will end India's dependence on foreign rockets for launching its remote sensing satellites; until now, India has been paying Russia \$30 million per launch. □

If you've ever developed a purification strategy by yourself, you know there is a lot to consider. Creating media screening schemes; designing buffer preparation routines; selecting which columns to use—even listing the tasks takes careful consideration. But now there's a better way of working.

Introducing ÄKTAexplorer

Turn ÄKTA™explorer on, choose a protocol, check running parameters and press start. ÄKTAexplorer does the rest.

Designed for all chromatographic techniques, this new purification system offers you pre-set protocols for every major purification task—including method scouting and media

screening. You'll save time as the system automatically recommends the best columns for your runs. You get fast pH screening as it automatically prepares your buffers from stock solutions. The moment you turn ÄKTAexplorer on, you're presented with a direct path to full-scale purification.

That path is UNICORN®—ÄKTAexplorer's control software. Your scale-up is simplified as the system's software can transfer and implement your methods on purification systems at all scales.

Of course, these are just a few of ÄKTAexplorer's features. So call us at 1 (800) 526 3593 from North America, +81 (0)3 3492 6949 from Japan or +46 (0)18 16 50 11 from the rest of the world (or meet us on the Internet at <http://www.biotech.pharmacia.se/akta.htm>).



Pharmacia Biotech
Uppsala, Sweden. (And the rest of the world)

New Zealand science system

SIR — In a recent Commentary article (*Nature* 379, 112; 1996), David Swinbanks reported on an address by Professor Philippa Black to a *Nature*-sponsored conference in Canberra. Among other things he claimed that Black gave a “stunning example of harm done to basic research through strategic planning”, and also referred to “horrendous complexity and bureaucracy”.

On the first point, it is somewhat ironical that *Science* published the following day (271, 141; 1996) an article by Elizabeth Pennisi headed “Fund fuels a resurgence of basic research”. It described how a new fund, soon to be worth NZ\$25 million a year, has been established by the New Zealand government specifically to foster basic curiosity-driven research.

Similarly, it is peculiar that *Nature* chose to labour issues relating to the ratio between basic and strategic science. A recent survey conducted by the New Zealand Association of Scientists showed that this ratio was not identified as being of particular concern to New Zealand scientists. And, whatever criticisms may be legitimately levelled at the New Zealand science system, the trend of government funding is upwards, unlike that in most OECD economies.

I regret to say that Black's paper contained many inaccuracies. My detailed letter of response listing 28 of the problems raised by her speech has been published on the same website as her address.

The science system in New Zealand is new and in need of continuing improvement and adjustment. But virtually none of the justified criticisms can be found in Black's speech. A working scientist who is by no means enamoured of every aspect of the new system commented, when I showed him her speech, that whatever criticisms he and his colleagues have, he did not recognize them in the speech. It struck him as being written by somebody who was not particularly close to the actual operation of the system. The level of simple factual inaccuracies and the absence of evidence to support Black's most serious charges renders her paper, in my view, almost useless as a commentary.

Finally, I believe *Nature* did not do justice to Black's paper, in which she took considerable trouble to spell out some of the benefits of change as well as the costs. No one reading your report could guess

that she ended her analysis by noting a “degree of optimism within the New Zealand science community that has not been seen for almost a decade”.

Simon Upton

(Minister of Research,
Science & Technology)
Parliament Buildings,
Wellington, New Zealand

David Swinbanks replies: Those attending the Canberra conference were taken aback by Philippa Black's description of the dramatic changes in the funding and management of New Zealand's science over the past ten years. It became the topic for much discussion among delegates, mainly along the lines of “at least things are not that bad in my country”. That is why we chose to focus on her presentation.

The reference to the “horrendous complexity and bureaucracy” of New Zealand's grant system paraphrased both the words and tone of Black's speech which variously described the system as “very complex and immensely bureaucratic” and “very bureaucratic and extremely conservative”.

Black's speech did end on a positive note, mentioning the Marsden Fund for basic research to which Simon Upton refers. But she pointed out that in 1995 it constituted only 2% of the Public Good Science Fund, and the main thrust of her speech was on the damage that she claims has been done over the past ten years to basic research in New Zealand as a result of government reforms. Readers can see Black's full speech and Simon Upton's rebuttal on the following website maintained by the Australian National University: <http://biology.anu.edu.au/Pages/Pubs/NatConf/Nathome.html>

Metallic optic fibres unlikely

SIR — Daedalus¹ proposes that metallic light-wires will act as optic fibres, transmitting light with a delay appropriate to the high refractive index characteristic of metals. Unfortunately, the refractive index of a typical metal is dominated by a term in $\sqrt{i}$ so that the propagation of a light wave is in fact highly damped in a metal — for example, in sodium² at room temperature, a wave of vacuum wavelength 10 μm will have a damping length of 0.1 μm . This may be contrasted with the 100-km damping length found for light in high-quality glass fibres. The optic fibre companies are not “likely to lose their monopoly of wide-band optical transmission”.

Charles Kittel

PO Box 368,
Philo, California 95466, USA

1. *Nature* 379, 124 (1996).
2. Kittel, C. *Introduction to Solid State Physics* 7th Edn, 331 (Wiley, New York, 1996).

Support fishermen and dolphins

SIR — Your recent article “Environmental lobby splits on US tuna bills” (*Nature* 379, 288; 1996) details two approaches to dealing with marine mammal interactions in the eastern Pacific purse seine tuna fishery.

Marine mammals are present in every ocean and therefore virtually every fishery. To put this fishery in perspective, the present US Marine Mammal Protection Act would consider a take of approximately 55,000 dolphins annually to be biologically insignificant on dolphin populations of the numbers present in the eastern Pacific. This is due to the huge size of this fishery (8 million square miles, almost 3 times the size of the United States) and the abundance of marine mammals (10 million). The international fishermen of the eastern Pacific are, however, committed to a higher standard by participating in the education programme of the Inter-American Tropical Tuna Commission. They are committed to reducing marine mammal deaths in the fishery to fewer than 5,000 animals by 1999. They reached and surpassed that target 2 years ago.

The continuation of the ‘dolphin safe’ status quo, as in Senator Barbara Boxer's bill, continues to oppose the encirclement and release of dolphins by purse seine net fishermen on moral grounds even though substitution of other gear would result in higher marine mammal mortality rates. Boxer's bill offers no support for the fishermen and what they have accomplished. However, Greenpeace, the Center for Marine Conservation, World Wildlife and other groups are supporting Senator John Breaux's bill, which will remove US embargoes from this fishery in the light of the tremendous progress made.

Breaux's bill does not remove or reduce regulation of the fishery — in fact, it adds more protection for stocks of dolphins. But it does expand the definition of ‘dolphin safe’ to include fish produced by purse seine fishermen when all the dolphins involved are released unharmed as verified by on-board observers. This allows the fishermen to catch clean schools of large yellowfin to keep the fishery healthy, and ‘dolphin safe’ becomes a well-earned ‘gold star’ for perfect performance.

We need to support effective multi-lateral conservation programmes and the efforts of responsible fishermen, not blacklist their products.

Teresa Platt

The Fishermen's Coalition,
826 Orange Avenue,
Apt 504,
Coronado,
California 92118, USA

Correspondence

Letters submitted for Correspondence should be typed, double-spaced, on one side of the paper only, or e-mailed to nature@nature.com

Dust from beyond the Solar System

David W. Hughes

A sensitive radar in New Zealand has for the first time detected the faint meteors produced by interstellar dust grains hitting the Earth. It even points to nearby, young stars as being among the sources.

DUST particles are continually striking the Earth's upper atmosphere. Each meteoroid rapidly loses mass and velocity, leaving behind it a train of excited and ionized gas. This train can be seen as a meteor (colloquially known as a shooting star), and can also reflect radar pulses (when it is known as a radio meteor).

Almost all meteoroids have incident velocities between 11 and 73 km s⁻¹. The former is the Earth's escape velocity, so anything falling to Earth will be travelling at least that fast. The latter velocity is reached in a head-on collision with a meteoroid in a parabolic orbit around the Sun; that is, one with just enough energy to be on the border between a bound, closed orbit and an unbound one. Meteoroids in the above velocity range are members of the Solar System, and have been produced by the decay of comets and by collisions between asteroids.

Throughout the history of meteor science, people have wondered whether any incident meteoroids have hyperbolic orbits (with enough energy to be unbound). Rather than being miniature planets of our own Sun, hyperbolic meteoroids would have an origin beyond the Solar System. The easiest way of detecting them is by looking for a geocentric velocity in excess of the upper limit of 73 km s⁻¹ given above. Öpik estimated¹ that between 60 and 70 per cent of the meteoroids that were not associated with the well-known showers were hyperbolic. Lovell² thought that this value was far too high, and by the 1960s Öpik^{3,4} had reduced his value to around 3 per cent.

It is clear that our Solar System must be a source of hyperbolic meteoroids for other planetary systems. Consider comet Swift-Tuttle, the parent of the Perseid meteor shower that is active throughout most of August. Swift-Tuttle has an orbital period of 130 years and at perihelion it has a velocity of 42.2 km s⁻¹, very close to the parabolic limit. At perihelion the 'dirty snowball' cometary nucleus is sublimating furiously, and millimetre and submillimetre dust particles are being blown away from its surface at a relative velocity of around 0.9 km s⁻¹. Any dust that is emitted in the direction of the comet's motion moves off on a hyperbolic orbit⁵, and typically 40 per cent of the dust lost by this comet leaves the Solar System and moves off through the Galactic disk.

The proportion of hyperbolic meteoroids produced increases as the meteoroid mass decreases. Comets with shorter orbital periods and with nuclei that have direct rather than retrograde spin will be considerably less prolific when it comes to producing hyperbolic meteoroids.

On page 323 of this issue, Taylor, Baggaley and Steel⁶ recount how they have used a powerful radar telescope near

fast-spinning A- and B-type stars are unlikely to have planetary systems because their original stellar nebulae are too hot for planets to condense⁷. Planetary systems with large numbers of gas giant planets, and thus decaying comets, are prevalent around single G and K stars with masses in the 1 to 0.25 solar mass range. Maybe the New Zealand high-velocity particles are produced by a differ-

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

The trail of a Perseid meteor at dusk. The dust particle that produced this meteor came from a comet within our Solar System, but tiny meteors from other stars have now been detected by radar. No visible shooting stars have yet been shown to be of interstellar origin.

Christchurch, New Zealand, to measure the orbits and sizes of incident meteoroids. Just under 1 per cent of all the meteoroids they detected had velocities in excess of 100 km s⁻¹. As this is far above the parabolic limit, the authors are confident that these meteoroids are from beyond the Solar System. The typical diameters of the detected meteoroids were between 15 and 40 µm.

What is even more exciting is that the annual variability of the hyperbolic meteoroid flux reveals the direction of the sources of these particles. Three major sources are identified. One is a group of nearby luminous A-type stars; another is a local group of even brighter B-type stars and young star clusters moving past the Sun; and the third is associated with the direction in which the Sun is moving around the Galaxy.

It is interesting to note that massive,

ent mechanism from the one seen in our own Solar System.

It also seems that the fraction of hyperbolic meteoroids decreases as the particle size goes up. Still, it is refreshing to know that even though more than 99 per cent of the dust in the Solar System is 'home grown', at least a fraction of a per cent comes from beyond. □

David W. Hughes is in the Department of Physics, University of Sheffield, Sheffield S3 7RH, UK.

1. Öpik, E. J. *Publ. astr. Obs. Tartu* **30**, No. 6 (1941).
2. Lovell, A. C. B. *Meteor Astronomy* 212-247 (Clarendon, Oxford, 1954).
3. Öpik, E. J. *Contr. Armagh Obs.* **26**, 1-82 (1958).
4. Öpik, E. J. *Jr. astr. J.* **9**, 156-159 (1969).
5. Harris, N. W., Yau, K. K. C. & Hughes, D. W. *Mon. Not. R. astr. Soc.* **273**, 999-1015 (1995).
6. Taylor, A. D., Baggaley, W. J. & Steel, D. I. *Nature* **380**, 323-325 (1996).
7. Hughes, D. W. *New Scientist* **136**, No. 1851, 29-33 (1992).

David Nunuk/SPL

Educating effector T cells

Massimo Trucco and Giorgio Stassi

ON page 352 of this issue¹, Duchosal and collaborators describe the use of a new treatment that abolishes the graft-versus-host reaction, as well as lymphoproliferative diseases, in mice transplanted with human blood cells. The human cells persisted in the mice, indicating long-term graft survival. These experiments outline a potentially effective approach towards regulating the immune system, in which a protease receptor, present on 5–10 per cent of resting T lymphocytes, is targeted.

The technical ability of surgeons to transplant various types of organs from, for instance, the victims of auto accidents to terminally ill patients has yet to be met with equal success in controlling the immune reactions involved in graft acceptance and survival. The conundrum of modern transplantation remains the need to overcome the anti-graft immune response without annihilating the entire immune system of the recipient. Severe suppression of the immune system is necessary to maintain graft survival, leaving recipients at high risk of developing graft-versus-host disease, rampant and potentially deadly infections, and unusual neoplasms such as B-cell lymphomas.

These post-transplant complications are particularly acute in patients receiving organs that are only partially 'matched' at relevant loci of the major histocompatibility complex; or, in the extreme cases in which animal organs are transplanted into human recipients (xenotransplants), where immunosuppressive regimes are, by necessity, extraordinarily powerful. Such transplants will be performed even more frequently as transplant centres attempt to provide an option to recipients for whom there is not a good 'match', as well as to meet the need for more organs than are currently being donated.

Transplantation of bone marrow presents particular obstacles. Among the donor bone marrow², and perhaps also among recipient bone marrow³, are subpopulations of cells that seem to facilitate or maintain engraftment⁴. Thus eliminating aggressive, activated T cells from the bone marrow of the donor in an attempt to avoid graft-versus-host disease may also contribute to poor engraftment of the foreign bone marrow. Similarly, because most bone-marrow recipients suffer from leukaemia or other neoplasms, their bone-marrow cells, including the cells that could contribute to maintaining engraftment, are killed by the robust irradiation treatment normally performed before transplant.

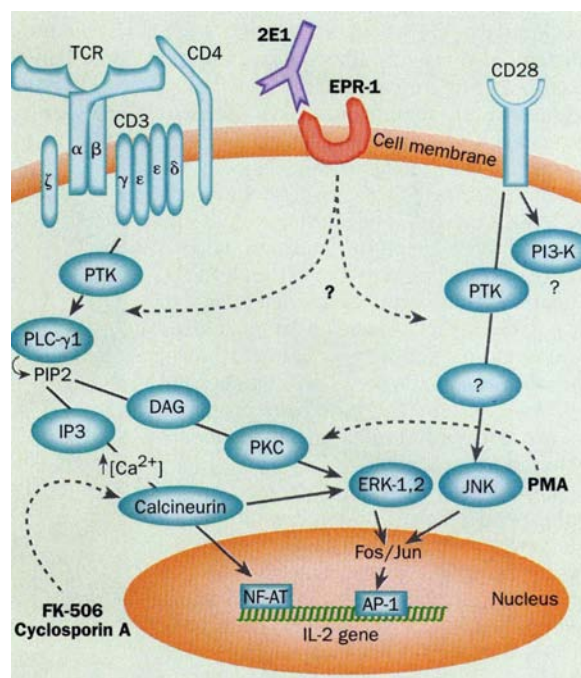
The effector cell protease receptor-1 (EPR-1) is a cell-membrane receptor involved in blood coagulation, signal

transduction, cell growth and, as the new study demonstrates, lymphocyte activation. Duchosal *et al.*¹ used a mouse monoclonal antibody (2E1) directed against human EPR-1 to modulate the donor–recipient immune interplay following xenotransplantation. In the presence of the 2E1 antibody, human peripheral blood lymphocytes (PBL) were successfully transplanted into severe combined immunodeficient (SCID) mice. Moreover, if PBL infected with Epstein–Barr virus were injected, mice receiving antibody treatment did not develop B-cell lymphomas as the non-treated controls did. The possibility of using monoclonal antibodies to mediate immunosuppression, to reduce the potential for graft-versus-host disease without removing precious cells from the transplant inoculum, as well as to prevent the development of B-cell lymphomas, is not only welcome but is actually quite provocative, particularly in the context of human organ transplantation.

Although the mechanisms by which the 2E1 antibody exerts its immunosuppressive activity are not known, it could be that it is its mild effect (in contrast to that of more powerful agents such as OKT3, cyclosporin A or Tacrolimus) that underlies the results achieved by Duchosal *et al.* (see figure). The lymphocytes affected by 2E1 are not directly killed or depleted *in vivo*, but are apparently only prevented from being fully activated. The authors provide evidence that the antibody may interfere with protease-dependent mechanisms of lymphocyte co-stimulation. Presumably, maintenance of the target cells in an inactive state can preserve the immunological donor–recipient cell balance in the absence of unwelcome side effects, although the potential results of 2E1 antibody treatment on host immune cells remain to be determined.

The activity of the 2E1 antibody treatment is actually reminiscent of (or perhaps mimics) that of cells which veto the activation of certain effector T cells without affecting other cells that can continue to perform their functions⁵. In the transplant recipient, cells with this type of regulatory activity also seem able to establish a chimaeric status in which the truce between recipient and donor immune systems guarantees the survival of the graft, without the need for protracted immunosuppression⁶.

If Duchosal and colleagues' experiments can be reproduced *in vivo* (in humans, for example, after the 2E1 mono-



Signalling pathways within a T cell, showing how the activity of the antibody 2E1 interferes with protease-dependent mechanisms of lymphocyte co-stimulation. Cell stimulation induced by T-cell receptor (TCR)/CD3 is mediated by a protein tyrosine kinase (PTK) which induces phosphorylation and then activation of a phospholipase $\text{C}\gamma 1$ ($\text{PLC-}\gamma 1$). In turn, $\text{PLC-}\gamma 1$ activation determines the hydrolysis of phosphatidylinositol-4,5-bisphosphate (PIP_2), and leads to the production of the second messengers inositol trisphosphate (IP_3) and diacylglycerol (DAG). These second messengers are responsible for mobilization of cytoplasmic free calcium and activation of protein kinase C (PKC), respectively. On the one hand, the Ca^{2+} increment determines the activation of transcription factors such as NFAT through calcineurin, and AP-1 through mitogen-activated protein kinase (MAPK) ERK-1,2 (ref. 7); these factors are responsible for the transcription of the interleukin (IL)-2 gene. On the other hand, PKC activates Ras which is involved in the transcriptional activation of Fos/Jun⁸. Although the real nature of CD28-mediated co-stimulatory transduction signal is not yet known, it is evident that the TCR/CD3 signal is up-regulated by the activated CD28 which directly influences MAPK-JNK⁹. The binding of monoclonal antibody 2E1 to EPR-1 down-regulates the effect of the CD3 and/or CD28 co-stimulatory signal, possibly inhibiting a PTK-dependent enzyme. Once bound to cytoplasmic proteins (immunophilins), cyclosporin A and FK-506 (Tacrolimus) form complexes with high affinity for calcineurin, thus inhibiting the transcription of several cytokines. Together with calcium ionophores, phorbol myristyl acetate (PMA) synergistically induces production of IL-2, acting directly on PKC. Interleukin-2 enhances T-cell activity, and so the 2E1 antibody has antiproliferative effects.

clonal antibody is 'humanized', or following the development of other antibodies with similar activities), we can envisage its use in treating donor bone-marrow cells before transplanting them. Antibody pretreatment would, one hopes, eliminate the need for the clinician to elutriate the donor bone marrow, because this process of eliminating activated T cells also tends to remove cells that are necessary for efficient engraftment of donor tissue. The monoclonal antibody treatment could be continued *in vivo* until the danger of developing graft-versus-host disease had passed, then discontinued. But it could be reinstituted promptly should a life-

threatening lymphoproliferative disorder present itself.

If such treatment works, a new era for tissue transplantation might begin — one in which more transplants would be performed, in which transplants would survive longer, and in which the recipient would be less exposed to the damaging consequences of collateral immune reactions. □

Massimo Trucco and Giorgio Stassi are in the University of Pittsburgh School of Medicine, Children's Hospital of Pittsburgh, Rangos Research Center, 3460 Fifth Avenue, Pittsburgh, Pennsylvania 15213, USA.

1. Duchosal, M. A., Rothermel, A. L., McConahay, P. J., Dixon, F. J. & Altieri, D. C. *Nature* **380**, 352–356 (1996).
2. Martin, D. R. & Miller, R. G. *J. exp. Med.* **170**, 679–690 (1989).
3. Petit, T. R. *et al. Blood* **84**, 3575–3583 (1994).
4. Starzl, T. E. *et al. Hosp. Pract.* **10**, 31–42 (1995).

5. Thomas, J. M. *et al. Clin. Transplant.* **8**, 195–203 (1994).
6. Burlingham, W. J. *et al. Transplantation* **59**, 1147–1155 (1995).
7. Li, W. *et al. Science* **271**, 1272–1276 (1996).
8. Fields, P. E., Gajewski, T. F. & Fitch, F. W. *Science* **271**, 1276–1278 (1996).

ECOLOGY

Mystery of moribund marram

Peter D. Moore

MOST plants grow less well under uncomfortable conditions such as drought or shade, or in extremes of heat. But the reverse seems to apply to some plants associated with the colonization of difficult habitats like sand dunes: they become increasingly moribund as conditions improve. A conventional explanation has been that other, more competitive species invade and are able to oust the pioneers by their aggressive growth under the improving circumstances of the habitat, but the truth is rarely as simple as that.

In the latest issue of the *Journal of Ecology*, L. R. Little and M. A. Maun¹ have examined a range of interacting factors in the dunes around the shores of Lake Huron, Canada, which seem to play a part in the demise of marram grass (*Ammophila breviligulata*), particularly the role of arbuscular mycorrhizal fungi in conferring some degree of protection from plant pathogens. The question of what kills the marram has attained some importance for ecologists in recent years, because this grass has achieved the status of a model species in the debate concerning what governs the rate and the direction of vegetation succession^{2,3}.

There are only two species of marram grass, the European and North African *A. arenaria* and the North American *A. breviligulata*, and both play an important role in dune building in coastal and inland

areas where sand is blown and accumulates. Their capacity for trapping mobile sands in their dense tussocks of vertical leaves and binding it with extensive and vigorously growing rhizomes gives the marram grasses considerable economic importance in land reclamation and

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Marram grass — a model species in studies of vegetation succession, and of considerable economic importance in land reclamation and protection.

protection, particularly around low-lying regions where dunes represent a vital element in coastal defence. The ecological significance of *Ammophila* in dune establishment was commented upon by Tansley⁴, who in 1911 noted the ability of this plant to withstand burial by sand and also the way in which it survives into the later, more stable stages of dune development, but becomes "less luxuriant than on the mobile sand, and its leaves lose their fresh green appearance". The demise of the marram in mature dunes

has become a focus of research attention for ecologists over the intervening years, for it may supply a clue to the forces that drive ecological succession.

The establishment of a plant species in dunes is often a consequence of 'facilitation' — the modification of the physical habitat by a preceding species in such a way that conditions are now suitable for the invader's establishment and survival. Marram is no exception; its relative sensitivity to salt-water flooding means that in maritime situations it usually arrives after the development of small, embryo dunes by a more salt-tolerant grass, *Elymus juncea*. The robust structure of marram creates a very different microclimate, facilitating the establishment of many other plant and animal species. But why should it decline just as conditions seem to have ameliorated? The explanations have proliferated over the years, and although many have an element of truth, none is fully satisfactory.

Competition from other plant species cannot be eliminated as an explanation, but it is not very convincing because moribund marram is often surrounded by a relatively low-growing, species-rich sward of bryophytes, lichens and small herbs — hardly, one would think, a match for marram. Grazing certainly takes its toll, especially where rabbits are abundant, and enclosure experiments usually enhance marram performance. The lack of burial by new supplies of mobile sand is certainly associated with its sad state, and ecologists

have variously explained this in terms of surface drought, lack of rhizome bud development, excess soil CO₂ or a lack of soil O₂. Most recently, attention has turned to soil pathogens, particularly fungi and nematodes⁵.

This explanation is based on the idea that the maturing soil of the later dunes bears increasingly dense populations of organisms that have the capacity to debilitate their hosts, which in the case of marram leads to the observed decrease in luxuriance. Work on *A. arenaria* by Van der Putten and his colleagues in the Netherlands⁶ showed that when marram was grown experi-

mentally in soil from different stages in dune development, its productivity was enhanced when the soil had previously been sterilized by gamma irradiation. So, if pathogens in the soil were destroyed in this way, marram resumed healthy growth. Burial, therefore, provides the plant with new, relatively sterile space in which roots can develop and escape the ravages of the pathogens. In order to ensure that differences in marram growth in this experiment were not due to chemical nutrient variations in the soils from

Jack Dermid/Oxford Scientific Films

the different stages in dune development, nutrients were added to all samples. But this could have disguised some other interactions in the soil community, particularly the activity and importance of arbuscular mycorrhizal fungi in marram's mineral nutrition and pathogen resistance. This work, in other words, examined the role of harmful organisms in the soil but left obscure the possible role of growth-enhancing ones.

Little and Maun¹ have investigated this problem by using sterilized soil for all their experiments and introducing to the test seedlings inoculations of nematodes and arbuscular mycorrhizal fungi, either singly or in combination. In addition, they

tested the physical effect of sand burial to a depth of five centimetres. Burial was invariably stimulatory in its effects, but so was the mycorrhizal inoculation; this suggests that although the arrival of new sand may provide a means of escape from the growing populations of nematodes (or pathogenic fungi), the presence of mycorrhizal infection of the roots gives them a degree of resistance to attack. Indeed, the authors claim that the mycorrhizal connection is more important than physical burial in determining the health of *Ammophila*.

The message that emerges is that the forces driving succession in this instance are being exerted by microorganisms in

the soil. As is inevitably the case in a story based on species replacement, the bad guys win in the end, but not without an unseen microbial struggle. □

Peter D. Moore is in the Division of Life Sciences, King's College, Campden Hill Road, London W8 7AH, UK.

1. Little, L. R. & Maun, M. A. *J. Ecol.* **84**, 1–7 (1996).
2. Tilman, D. *Resource Competition and Community Structure* (Princeton Univ. Press, 1982).
3. Crawley, M. J. *Nature* **362**, 17–18 (1993).
4. Tansley, A. G. *Types of British Vegetation* (Cambridge Univ. Press, 1911).
5. Van der Putten, W. H. in *Primary Succession on Land* (eds Miles, J. & Walton, D. W. H.) 273–281 (Blackwell Scientific, Oxford, 1993).
6. Van der Putten, W. H., Van Dijk, C. & Peters, B. A. M. *Nature* **362**, 53–56 (1993).

OCEANOGRAPHY

Awakenings in the Arctic

Robie Macdonald

UNTIL only a few years ago, the Arctic Ocean was considered to be a quiet, unproductive backwater, hardly changing or causing change in the oceans to the south. This mistaken view was due partly to the sparseness of observations in the interior ocean, inaccessible because of the ice, but there seems also to be a consensus emerging that a change has recently occurred in the Arctic Ocean's circulation. A layer of water that derives from the Atlantic is warming up, and the exchange between the Arctic Ocean's two major basins — Canadian and Eurasian — has somehow altered.

At the Ocean Sciences meeting in February*, a special session organized by K. Aagaard (Univ. Washington) illustrated a coincidence in the arrival of change and the availability of some extraordinary tools to observe it. The use of icebreakers and submarines in the past few years, starting with the cruise of the *Polarstern* in 1987, combined with data from deep-sea moorings, have not only changed our perceptions of the Arctic Ocean but have also allowed us to see change occur. Icebreakers have created a time series of high-quality geochemical and biological sections through the ocean across regions of permanent ice cover, and submarines have produced large-scale, synoptic surveys undreamed of a decade ago.

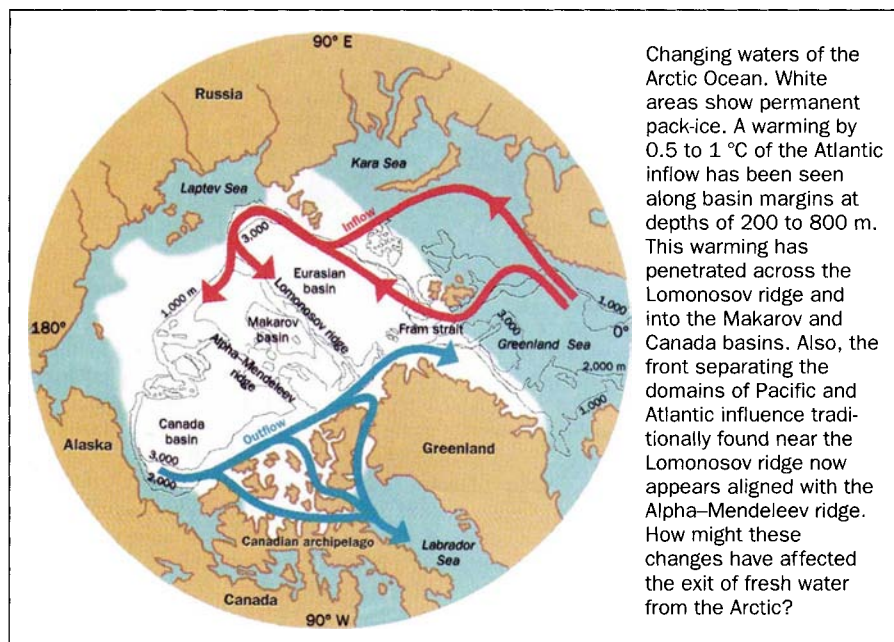
The observed warming and shallowing of the Atlantic layer (the region between about 200 and 800 m) was first reported¹ in 1993. The progression of the warming is confirmed by submarine data (J. H. Morrison, Univ. Washington) and by the Arctic Ocean Section made by two icebreakers in 1994 (J. H. Swift, Scripps). The core of the Atlantic layer is now 0.5 to 1 °C warmer than in any historical observations, which

go back to the 1950s, and this warming now extends well into the Canada basin. No one has suggested it to be a sign of climate change; rather it illustrates that the Arctic Ocean is dynamically coupled to the Atlantic and much more variable than previously suspected.

The invasion of warm water, seen by icebreakers as boundary currents around the Eurasian and Canadian basins, is accompanied by a shift of the front separating Atlantic and Pacific water masses in the upper 200 m of the Arctic Ocean. Historically, this front has been locked to the Lomonosov ridge, but it appears now to be over the Alpha-Mendelev ridge (see figure). Clearly, a large quantity of Atlantic water has displaced the Pacific water, insinuating itself into the Makarov basin as a front with the Atlantic core sloping down

from about 200 m in the Makarov basin to about 450 m in the Canada basin. The invasion is accompanied by interleaving between the warmer, invading water and the cooler, resident water (E. C. Carmack, Inst. Ocean Sciences). These interleaving features are thin (40 m) but coherent over at least 1,000 km, and they are important sites of thermodynamic and double-diffusive mixing, caused by the more rapid diffusion of heat than salt.

The transport of water from the borders of the Kara and Laptev seas into the Makarov and Canada basins is of great interest to those concerned with where radionuclides dumped on the Russian shelves might go². It is ironic, therefore, that the most compelling geochemical evidence of recent transit of Atlantic water along the Kara/Laptev shelf edge and across the Lomonosov ridge is to be found in radionuclides derived not from Russia but from European reprocessing plants (J. N. Smith, K. M. Ellis and L. Kilius, Bedford Inst.). The ¹²⁹I/¹³⁷Cs ratio in discharges from these plants jumped by several orders



*Changing Perceptions of the Oceanography of the Arctic Ocean, Ocean Sciences Meeting, San Diego, California, 12–16 February 1996.

of magnitude during the early 1970s, and continued to climb from the mid-1980s to the present, offering one of the best opportunities to track Atlantic water masses as they cross the Arctic Ocean.

Where has the displaced Pacific water gone? This Canada basin surface water contains a large proportion of the fresh water of the Arctic Ocean³. It can escape via the Fram strait into convective regions of the Greenland Sea or via the Canadian archipelago into convective regions of the Labrador Sea (Carmack). Clearly, the displacement of this surface water from the Arctic Ocean is of the utmost importance for global ocean circulation and therefore for global climate.

Biological oceanographers have long considered the interior Arctic Ocean to be of low productivity, owing to its short summer and ice cover⁴. That view has changed. Chlorophyll levels and biological productivities are much higher than previously reported, perhaps by as much as a factor of ten (P. A. Wheeler, Oregon State Univ.; M. Gosselin, Univ. Quebec). Much of the production is recycled and occurs close to or within the ice, which may ease the entry and concentration of airborne contaminants into the Arctic marine food web.

Two tasks have been set recently for those who would model the Arctic Ocean: to predict movement of contaminants and to predict the response of the system to global change. Numerical models with 'eddy-resolving' grid sizes and a better representation of ocean-bottom topography have been successful for other oceans, so vast improvements may be just around the corner for the Arctic (A. J. Semtner, Naval Post Graduate School, Monterey). An important step towards reality in Arctic Ocean models will be the inclusion of the Canadian archipelago as an outlet for surface water. It is difficult to see how global change can be addressed otherwise, as this is a crucial conduit both for fresh water and contaminants.

Where do we go from here? The Arctic Ocean has recently been perturbed by both natural and anthropogenic factors, with changing temperature and concentrations of chemicals and radioisotopes. We must grasp the unprecedented opportunity to follow these perturbations as they cascade through the Arctic Ocean and to learn how this ocean is coupled to other oceans in the context of change. When a dragon wakes, it is a bad time to sleep. □

Robie Macdonald is at the Institute of Ocean Sciences, PO Box 6000, Sidney, British Columbia, Canada V8L 4BZ.

The core curriculum

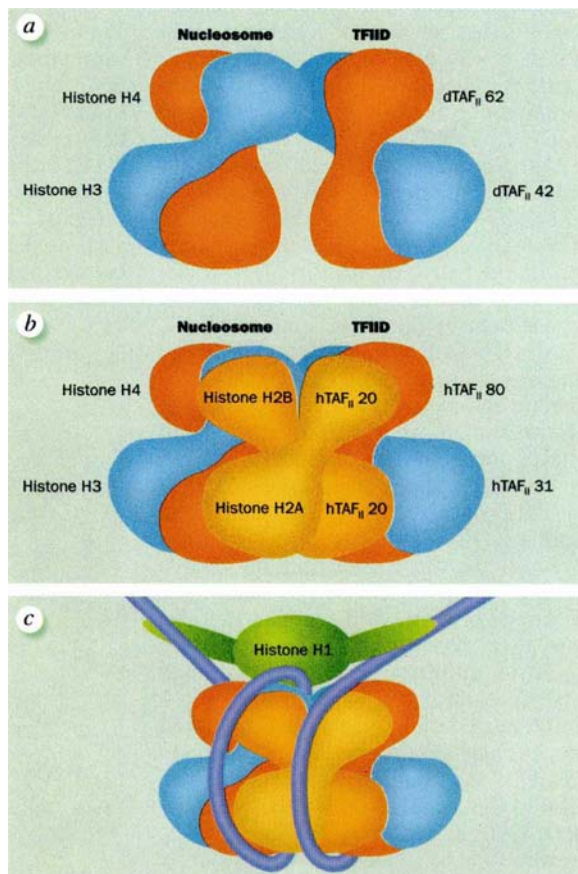
Christopher Surridge

In eukaryotes, some genes are copied into RNA much faster than others. The bulk of this control is exerted by proteins which bind to the DNA in and around the gene, affecting the process of transcriptional initiation. Although most steps in initiation are common to all protein-coding genes, involving a number of multi-subunit protein complexes, its details are still poorly understood. On pages 316 and 356 of this issue, Xie *et al.*¹ and Hoffmann *et al.*² further define the limits of our knowledge. Their results show that the transcription factor at the heart of the initiation complex, TFIID, contains a structure closely resembling that of the histones, proteins responsible for packaging the genome. Might the DNA at the promoter be bent or wound around TFIID in the same way as around the histone octamer?

Eukaryotes have three RNA polymerases, the second of which (Pol II) is responsible for transcribing protein-coding genes (reviewed in ref. 3). Pol II transcription begins when a sequence at the core of the promoter termed the TATA box is recognized by the TATA-box binding protein (TBP), a component of the multi-subunit complex TFIID. TFIID is a target for a wide range of transcriptional activators, and its binding depends on their activity; it also serves to direct subsequent steps in the initiation process. This involves the ordered binding of six further multi-subunit complexes, including Pol II itself. Following assembly of a productive initiation complex, details of which are unclear, DNA at the start of the gene is melted, allowing Pol II to bind to its strand and begin transcription. Pol II and some other parts of the complex are then released, but the remainder (including TFIID) stays attached to the promoter where it can support initiation by further Pol II molecules. These processes are controlled by various protein-protein interactions within the complex, which is thought to undergo major conformational changes. The three-dimensional structure

of the various parts of the complex and the way in which they fit together are thus of great importance in understanding how it operates.

Xie and co-workers have shed some light on this by using X-ray crystallography to determine the structure of a complex containing two TFIID components from *Drosophila*, the TBP-associated fac-



The structural and biochemical studies of Xie *et al.* and Hoffmann *et al.* point to similarities between the organization of histones and TAFs, shown here schematically. *a*, The shared arrangement of the H3/H4 histone and amino-terminal portions of the dTAFII42/dTAFII62 tetramer¹. *b*, The proposed arrangement of the hTAFII20/hTAFII31/hTAFII80 octamer² by analogy with the H2A/H2B/H3/H4 nucleosomal particle. *c*, The arrangement of the nucleosome, including core octamer, DNA and histone H1.

tors (TAFs) dTAFII42 and dTAFII62. The amino-terminal sequences of these subunits distantly resemble those of histones H3 and H4, respectively⁴, proteins which form the core of the nucleosomes used to package the cell's DNA. It was nevertheless surprising to find that they adopt the same protein fold, with a long central α -helix flanked at both ends by shorter helices. Even more surprisingly, both in crystals and in solution the two subunits form a tight heterotetramer, closely resembling the H3/H4 tetramer

1. Carmack, E. C. *et al.* *Geophys. Res. Lett.* **22**, 1061-1064 (1995).
2. Schlosser, P. *et al.* *Deep Sea Res.* **42**, 1341-1367 (1995).
3. Aagaard, K. & Carmack, E. C. *J. geophys. Res.* **94**, 14485-14498 (1989).
4. Subba Rao, D. V. & Platt, T. *Polar Biol.* **3**, 191-201 (1984).

at the core of the nucleosome⁵ (*a* in the figure).

To see why this result is so provocative, it is necessary to consider the normal function of the histones. Each human cell contains over a metre of DNA, which must be densely packaged if it is to be fitted into a nucleus with a diameter of a few micrometres. The basic packaging unit is an octamer formed from two molecules each of histones H2A, H2B, H3 and H4, around which 147 base pairs of DNA are wound in two-and-a-half turns which are 'sealed' in place by histone H1 (*c* in the figure). A linear array of such nucleosomes is then wound into a fibre 30 nm in diameter, which is itself coiled further to form chromosomes. Does TFIID also contain equivalents to histones H2A and H2B?

Using human TAF_{II}s, Hoffmann *et al.*² show that it does. The sequence of hTAF_{II}20 distantly resembles that of H2B, and binding studies showed that it can interact strongly with itself, with histones H3 and H4, and with the human equivalents of dTAF_{II}42 and dTAF_{II}62, hTAF_{II}31 and hTAF_{II}80. Together, these findings suggest that TFIID contains a histone octamer-like structure composed of two dimers of hTAF_{II}20 attached to a tetramer of hTAF_{II}31 and hTAF_{II}80 (*b* in the figure), a conclusion consistent with the observed stoichiometry of the complex.

What are the implications of such a finding for transcription initiation? Although proteins with similar structures can have very different functions, the TAFs in question retain the positively charged amino acids which contact DNA in the nucleosome⁶, suggesting that DNA may perhaps be wrapped around parts of TFIID in the same way that it is wrapped around the nucleosome. TFIID can compete with nucleosomes for DNA binding (reviewed in ref. 7), and confers a similar pattern of protection against cleavage by nucleases. Moreover, the TAFs which make up the putative octamer interact with several factors, including the tumour suppressor p53 (ref. 8) as well as other TAFs. If these interactions affect their capacity to bind DNA, they could trigger a conformational change in TFIID, as is thought to occur following binding of the transcriptional activator GAL4 (ref. 9).

As yet, the evidence that the histone-like arrangement does anything other than support protein-protein interactions is very circumstantial. But the possibility that it can bind DNA, perhaps in a controlled way during transcription initiation, is clearly exciting. Whatever the truth of

the matter, it is clear that the existence of such a structure will provide a fertile source of hypotheses, and experiments to test them. □

Christopher Surridge is an assistant editor at Nature.

ARCHAEOLOGY

Tool hafting with a mastic

Simon Holdaway

THERE is a growing body of evidence that some stone artefacts manufactured during the Middle Palaeolithic, a period conventionally dated to more than 35,000 years before the present (BP), were attached to a haft or handle. This is a technique used in more recent times to produce a wide range of complex, multi-component tools such as dart tips, barbed projectiles and cutting tools¹.

On page 336 of this issue² Boëda and

tic. Chemical analysis indicates that the bitumen was heated, presumably as part of the process whereby the tool was glued to the haft.

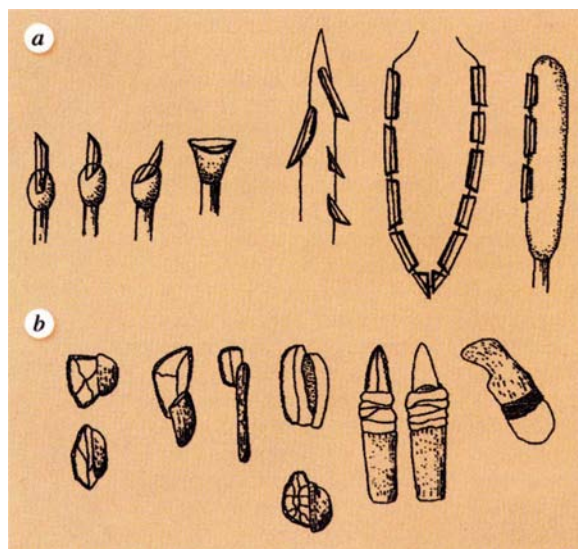
Stone is a relatively intractable medium in comparison to, say, clay or wood. Similarly shaped stone artefacts occur in many places around the world at many different times because the same solutions for working stone were repeatedly invented. This creates problems for archaeologists

who attempt to use similarity in the shape of artefacts to identify regions occupied by related peoples during particular periods. This problem can be overcome somewhat if artefacts can be shown to be hafted. Attaching a stone artefact to a handle often involves careful shaping of the stone to ensure a secure fit. This means that the stone artefact may acquire a more distinctive form, and so be more useful for identifying regularities in time and space.

Hafting is also important for assessing the pace of technological development. Middle Palaeolithic or Mousterian (the terms are used interchangeably to refer to the period before 35,000 years BP in Europe and the Near East) artefacts appear to be less

complex than those belonging to the Upper Palaeolithic (dating after 35,000 BP to 10,000 BP). But evidence for hafting in the Middle Palaeolithic may indicate that more complex multi-component forms existed earlier, so changing our perceptions of the relationships between the two periods.

Previous studies of Middle Palaeolithic artefacts have used traces of surface wear as evidence of whether a tool was hafted or not^{3,4}, and it has been suggested that differential patterns of patina (a colour change on the surface of stone as a result of weathering) on an artefact of that period also indicates that it was hafted⁵. The



Hafted stone artefacts. *a*, Upper Palaeolithic composite points, barbs and knives (redrawn from ref. 1). *b*, Examples of implements used by present-day peoples and which are of a form similar to artefacts found in Middle Palaeolithic contexts. They illustrate a range of ways in which Middle Palaeolithic artefacts may have been hafted (redrawn from ref. 3).

his colleagues present the results of a study that demonstrates the presence of traces of bitumen, interpreted as evidence for hafting, on two stone artefacts dating to this period, from the site of Umm el Tlel in Syria (see map, page 336). Both of the artefacts are pointed flakes manufactured according to a specialized method for flaking a core (called the Levallois technique) and one of them has its edges modified through retouch so that they meet to form a point (and thus a 'convergent' scraper). The bitumen appears to have been applied to the proximal region of the flakes, suggesting that they were hafted with the points free from the mas-

1. Xie, X. *et al.* *Nature* **380**, 316–322 (1996).

2. Hoffmann, A. *et al.* *Nature* **380**, 356–359 (1996).

3. Hori, R. & Carey, M. *Curr. opin. Genet. Dev.* **4**, 236–244 (1994).

4. Kokubo, T. *et al.* *Nature* **367**, 484–487 (1994).

5. Arents, G., Burlingame, R. W., Wang, B.-C., Love, W. E. & Moudrianakis, E. N. *Proc. natn. Acad. Sci. U.S.A.* **88**, 10148–10152 (1991).

6. Arents, G. & Moudrianakis, E. N. *Proc. natn. Acad. Sci. U.S.A.* **90**, 10489–10493 (1993).

7. Owen-Hughes, T. & Workman, J. L. *Crit. Rev. Euk. Gene Express.* **4**, 403–441 (1994).

8. Thut, C. J., Chen, J.-L., Klemm, R. & Tijan, R. *Science* **267**, 100–104 (1995).

9. Horikoshi, M., Carey, M. F., Kakidani, H. & Roeder, R. G. *Cell* **54**, 665–669 (1988).

examples from Umm el Tlel, however, seem to be the earliest so far discovered with direct evidence of a mastic. In the absence of Middle Palaeolithic artefacts still attached to the haft, the presence of a mastic provides the most convincing evidence to date that Middle Palaeolithic artefacts were attached to handles.

The claim that the two artefacts are Middle Palaeolithic in age is based on thermoluminescence dates and accelerator mass spectrometer radiocarbon determinations on samples from the layer immediately above those containing the artefacts. These techniques returned results of $36,000 \pm 2,500$ and $34,530 \pm 890$ respectively (the thermoluminescence dates are in calendar years; the spectrometer dates in radiocarbon years uncorrected for secular variation). Thus, the two artefacts with traces of mastic must be at least 36,000 years old. Boëda and colleagues' suggestion that they are Middle Palaeolithic is based on the possibility that they are much older, and on their form, which is characteristic of other artefacts from the same region and period.

If the artefacts are confirmed to be Middle Palaeolithic, two issues arise. First, previous work indicates that Middle Palaeolithic scrapers such as the retouched example identified from Umm el Tlel owe much of their shape and size to a process of repeated edge reworking or resharpening⁶. It has also been suggested that the majority of the variability in the proportions of retouched tools found in many Middle Palaeolithic assemblages can be accounted for by a model of tool reworking. This situation contrasts with the Upper Palaeolithic where fewer tool forms appear to have been regularly resharpened. In light of this, it will be interesting to discover whether hafted and non-hafted Middle Palaeolithic scrapers were resharpened in different ways and whether this difference is reflected in their shape.

Second, the new finds add to what is an increasingly complex picture of the Middle and Upper Palaeolithic in western Asia. At present, there is no clear-cut association between one hominid species and the Middle Palaeolithic and no clear notion of the date at which this period gave way to the Upper Palaeolithic. Hominid finds at the sites of Qafzeh and Kebara (Israel), dated by thermoluminescence of burnt flint, show both Neanderthal and anatomically modern individuals associated with assemblages of Middle-Palaeolithic-like stone artefacts, the archaic species having a more recent date than the anatomically modern individuals⁷.

So there is no simple evolutionary sequence of archaic hominids making simple tools being replaced by anatomically modern hominids making complex tools. The finds from Umm el Tlel imply that

tools with relatively simple morphologies may in fact have been hafted into more complex composite tools. If they are much older than 36,000 years BP, then they make the existence of a simple/complex division in technology between the Middle and Upper Palaeolithic even less likely, because of the implication that the people who fashioned the Umm el Tlel tools may have been capable of producing composite tools similar in complexity to those used later on.

Alternatively, if the Umm el Tlel artefacts date close to 36,000 years BP, then they might belong to a transitional assemblage. In this case, they could conceivably have been made by members of an archaic population — perhaps by copying peoples making more complex tool forms. Or they might represent artefacts which have a shape akin to those from the Middle Palaeolithic and which were made for a particular purpose by a group of people who normally made more complex stone artefacts. These and other possibilities are all conceivable explanations based on the

CONDENSED MATTER

Two waters and no ice please

R. J. Speedy

EVERY few years measurements on supercooled water are extended to lower temperatures and ideas about our essential liquid are put to the test. Tanaka's paper¹ on page 328 of this issue extends earlier simulations of cold water by Poole *et al.*² to longer timescales, and argues that the liquid approaches a low temperature stability limit, where it transforms suddenly into a different liquid.

The van der Waals description of the liquid-gas transition provides the most familiar framework for understanding phase transitions and stability limits. It shows that the line in the pressure-temperature diagram dividing liquid and vapour ends at a critical point, where the liquid and gas become identical, and that at any lower pressure there is an upper limit on the temperature to which a liquid can be superheated. The locus of those limits is called a spinodal line. As the liquid approaches that line its heat capacity, expansivity and compressibility diverge, and boiling becomes inevitable.

When liquid water is cooled below its freezing temperature, its heat capacity, expansivity and compressibility vary dramatically^{3,4}. One way to explain such anomalous behaviour is to suppose that there is another spinodal that delineates a lower temperature limit on the stability of water at a temperature T_s of about 227 K (at one atmosphere), where it would presumably freeze. Until 1980 that idea was consistent with observations that even

information provided by Boëda *et al.*

Determining which is the more likely must await the publication of more data. In truth, any News and Views article dealing with the prehistory of this period and place could end on that note. But it is the very uncertainty of explanation that will continue to make the subject of compelling interest. □

Simon Holdaway is in the School of Archaeology, La Trobe University, Bundoora 3083, Australia.

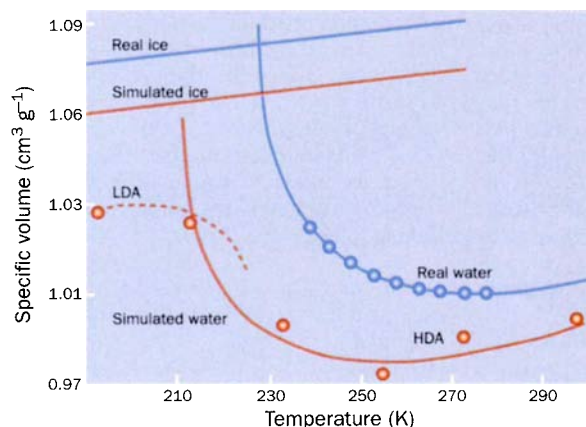
1. Bar-Yosef, O. in *La main et l'outil: manches et emmanchements préhistoriques* (ed. Stordeur, D.) 155–164 (Travaux de la Maison de l'Orient, Lyon, 1987).
2. Boëda, E. *et al.* *Nature* **380**, 336–338 (1996).
3. Anderson-Gerfaud, P. & Helmer, D. in *La main et l'outil: manches et emmanchements préhistoriques* (ed. Stordeur, D.) 37–54 (Travaux de la Maison de l'Orient, Lyon, 1987).
4. Beyries, S. in *La main et l'outil: manches et emmanchements préhistoriques* (ed. Stordeur, D.) 55–62 (Travaux de la Maison de l'Orient, Lyon, 1987).
5. Friedman, E. *et al.* *J. Israel Prehist. Soc.* **26**, 8–31 (1994–95).
6. Dibble, H. *Am. Antiq.* **52**, 109–117 (1987).
7. Bar-Yosef, O. *et al.* *Curr. Anthropol.* **33**, 497–550 (1992).

microscopic water samples always froze³ when cooled below 232 K.

However, Brüggeller and Mayer⁵ showed in 1980 that micrometre-sized drops did not freeze when cooled to the temperature of liquid nitrogen (77 K) at rates close to a million degrees per second. In 1994 Bartell and Huang⁶ cooled drops containing about 6,000 molecules from 235 K to 200 K in 40 microseconds by evaporation, and monitored the electron diffraction pattern. They did not see crystals until the cooling rate levelled out near 200 K. These experiments poured cold water on the notion that the liquid must freeze near 227 K.

A possible interpretation is that the liquid becomes so viscous near 227 K that it had no time to freeze in those fast experiments. However, it is precisely the forte of computer simulation that it can observe small systems in great detail on nanosecond timescales, and Tanaka¹ finds that the coldest sample studied, at 193 K, is a liquid that equilibrates in a few nanoseconds. This establishes that the falsification of the idea that water must freeze near 227 K cannot be dismissed by supposing that the water is structurally arrested on the microsecond timescale of the cooling experiments.

An apparently new form of water was made in 1984 by compressing ice to 10 kbar at 77 K, where it 'melts' from the usual crystalline state into a high-density amorphous solid^{7,8}, or HDA. The pressure



Plots of specific volume against temperature for simulated water and ice (red, from Fig. 1 on page 328 of this issue) and real water and ice (blue, from ref. 10), at atmospheric pressure. Experimental points for real supercooled water lie on a curve (equation 12 of ref. 10) whose slope diverges at the spinodal temperature $T_s = 227$ K. The same curve describes the simulation results if T_s is shifted from 227 to 211 K and the volume at T_s is reduced by 3 per cent. The dashed line is a guess at how the volume of the new phase, LDA, may vary. The guess is guided by Tanaka's phase diagram on page 329, which requires that the slope of the LDA line diverge to minus infinity at a temperature, T_s , above the T_s of HDA.

of this transition can be found by extrapolating the equilibrium ice–water melting line to 77 K, which suggests that HDA, although a glass, belongs to the same phase as ordinary water. But when HDA is taken to low pressure and then warmed up to about 120 K, it transforms into a low density amorphous solid (LDA), with a 20 per cent increase in volume^{7–9}.

The startling notion here is that there can be more than one distinct amorphous phase of the same condensed substance, separated by a sharp transition. That is a seminal discovery. The suggestion that HDA might transform continuously to ordinary water if warmed from 77 K to room temperature at high pressure has not been tested experimentally because all the amorphous phases of water freeze quickly at intermediate temperatures.

Sharp transitions between phases of different symmetry, like the melting of a crystal, do not have critical points, because the two phases cannot become identical. However, amorphous phases (gases, liquids and glasses) have the same symmetry, and transitions between them can, like the liquid–gas transition, end at a critical point. Poole *et al.*² exploited that possibility to explain cold water's anomalies by proposing an HDA–LDA transition, with a critical point and associated spinodal lines on either side of the coexistence line. That work², Mishima's^{7–9} and Tanaka's suggest that liquid water can transform to a distinctly different amorphous phase, and Tanaka now reports that this is just what happens in cold water. The anomalies of supercooled water are still attributed to a spinodal instability, but the unstable liquid (which belongs to

the HDA phase) is no longer required to freeze near its stability limit because it can transform into LDA, which turns out to be a liquid, rather than into ice.

The most familiar anomaly of water, that it expands when cooled below 277 K, and its explanation in terms of a spinodal, are illustrated in the figure. The van der Waals theory, and any other analytical theory, predicts the form¹⁰ of the relationship between volume and temperature for liquids near a spinodal shown in the figure. The slopes diverge as the temperature T approaches the spinodal temperature T_s , and that requires the observed increase in volume as the liquid is cooled.

Note that the volume at 193 K is far away from the HDA line, and so the

guessed dashed line that I have drawn to represent the LDA phase shows that somewhere near T_s the volume, or its temperature dependence, changes sharply. The lines drawn for HDA and LDA are consistent with Tanaka's results and with his proposed phase diagram, but show that the simulated densities can be explained either by an HDA to LDA phase transition between 213 and 233 K or by the presence of a transition not far below 213 K.

It is implicit in Tanaka's scheme that LDA is more stable than HDA above T_s . It might not transform back to HDA until near the high-temperature end of the dashed line shown in the figure, so properties measured on warming LDA from 195 K should differ from those measured on cooling HDA. Direct observation of that hysteresis would provide compelling evidence for the two distinct phases of liquid water. □

R. J. Speedy is in the Department of Chemistry, Victoria University of Wellington, PO Box 600, Wellington, New Zealand.

1. Tanaka, H. *Nature* **380**, 328–330 (1996).
2. Poole, P. H., Sciortino, F., Essmann, U. & Stanley, H. E. *Nature* **360**, 324–328 (1992).
3. Angell, C. A. in *Water: A Comprehensive Treatise* vol. 7 (ed. Franks, F.) 1–82 (Plenum, New York, 1982).
4. Speedy, R. J. & Angell, C. A. *J. chem. Phys.* **65**, 851–858 (1976).
5. Brüggeller, P. & Mayer, E. *Nature* **288**, 569–571 (1980).
6. Bartell, L. S. & Huang, J. *J. phys. Chem., Ithaca* **98**, 7455–7457 (1994).
7. Mishima, O., Calvert, L. D. & Whalley, E. *Nature* **310**, 393–395 (1984).
8. Mishima, O., Calvert, L. D. & Whalley, E. *Nature* **314**, 76–78 (1985).
9. Mishima, O. *J. chem. Phys.* **100**, 5910–5912 (1993).
10. Speedy, R. J. *J. phys. Chem., Ithaca* **91**, 3354–3358 (1987).

National expert

DEMOCRACY these days depends heavily on opinion polls. Existing polls are mere brief snapshots of public opinion on a narrow range of issues. Daedalus wants to improve them. His new polling process will deduce the popular opinion on every possible social or political topic.

It is based on those computer 'expert systems' which codify the judgement of one human expert into many thousands of simple rules. The best of them can mimic their 'role model' fairly plausibly over a narrow range of topics. Daedalus is devising a mighty Democratic Expert System, DEX, to encapsulate the feelings and judgements of a whole nation.

The principle is very simple. Pollsters will travel or telephone round the land, asking thousands of people thousands of questions on all imaginable social and political topics. Not everybody need be polled on all issues. By standard sampling theory, even a few thousand responses to a question give a very good estimate of the spread of opinion on the topic. The responses will gradually be built up into a huge database.

But this is only the start. A conventional expert system would attempt to extract from the data a vast set of rules, from which the public response to any possible question could be deduced. This may not work. Any wide range of popular opinions will include numerous contradictions; its only consistency will be statistical. So Daedalus will consider DEX's data as a set of known points in some multi-dimensional 'issue space' (the Dewey universal decimal classification of knowledge is a simple one-dimensional issue space). Every possible topic will have its point in issue space, associated with the spread of opinion on that topic. By smooth interpolation between the known polled points, DEX will construct an 'opinion hypersurface' mapping the spread of opinion on all possible topics. It will then be able to predict the spread of popular opinion even on questions never asked by its pollsters.

DEX will be updated continuously, to keep it abreast of public opinion. Some opinions will be highly stable; others will vary rapidly, blown here and there by transient fashions, disasters or scandals. DEX will note this volatility, too. It will tell us and the politicians far more about our feelings than they get from one vote per person per few years. At any time, they will know just how we would view any proposed move on their part. They will be warned in advance of what they can and can't get away with. Disasters such as the UK's 'Poll Tax' fiasco will be nipped in the bud. With detailed feedback by DEX, democracy will work at last. David Jones

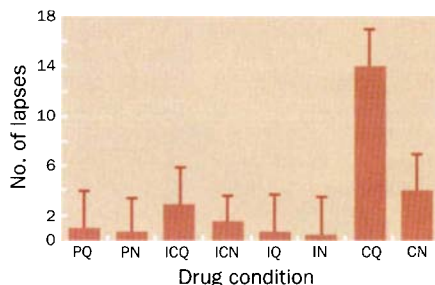
Noradrenaline and attention lapses

SIR — States of low alertness, produced by factors such as sleep deprivation, decrease performance on tasks of sustained attention. This is manifest by increases in the number of very long reaction times or in failures to respond. The effect can be reduced by increasing alertness by, for example, exposing the person to loud noise¹.

Although arousal is probably regulated by several neurotransmitter systems, noradrenaline is one of the most important. Brain noradrenaline is involved in the control of attention²; recently, it has been demonstrated that activation of noradrenergic neurons facilitates behavioural responses to subsequent sensory cues³. Noradrenergic neurons are particularly active during high states of arousal, such as those induced by exposure to noise⁴.

The clonidine challenge is an established method of manipulating brain noradrenaline function in humans⁵. Clonidine, an α -2-adrenoceptor agonist, in low doses acts presynaptically to decrease noradrenergic cell firing and noradrenaline release; these effects can be blocked by idazoxan, a selective α -2-adrenoceptor antagonist⁶. We predicted that lapses of attention would increase following clonidine challenge, and that this effect would be reversed either by exposure to noise or by co-administration of idazoxan.

We conducted a parallel-groups, double-blind study involving challenge with placebo, idazoxan, clonidine and idazoxan + clonidine. Half the subjects were tested in quiet and half were played broadband noise over headphones. Subjects were given a medical examination, a familiarization session, and then a baseline test followed by three post-drug tests. The study was approved by the local ethics committee and conducted with the informed consent of the subjects, 74 males, aged 18–35 years. Their task was a two-choice reaction time task involving focused attention and selective response to the letters A or B (ref. 7), which was



Effects of the different drug and noise conditions on lapses of attention (reaction times > 1,500 ms). Scores are adjusted means from an analysis of covariance. Bars, s.e. Drugs were given orally: 40 mg idazoxan; 200 μ g clonidine; the placebo was lactose. The noise levels were 75 dBA in the noise condition and 50 dBA in the quiet condition. Abbreviations for the various conditions: N, noise; Q, quiet; P, placebo; I, idazoxan; C, clonidine.

adapted to record lapses of attention also (defined as RTs > 1,500 ms). Baseline testing occurred between 09.00 and 10.30, administration of drugs at 10.30, and post-drug sessions at 11.00, 13.00 and 15.00.

The figure shows that the number of lapses of attention was markedly greatest in the 'clonidine, quiet' condition ($P < 0.05$). This effect was reversed by idazoxan and noise, and specific statistical comparisons showed that the clonidine, quiet mean was significantly greater than all the other conditions, which did not differ from one another.

These results clearly demonstrate that central noradrenaline is important in maintaining attention. The effect of clonidine was obtained using a dose that acts selectively at presynaptic receptors to inhibit noradrenaline release⁸, and was fully reversed by idazoxan, an antagonist of these receptors. The reversal of the clonidine effect by noise is consistent with the view that this stimulus increases arousal by increasing noradrenaline release.

Andrew Smith

David Nutt

Departments of Psychology and Psychopharmacology,
University of Bristol,
8 Woodland Road,
Bristol BS8 1TN, UK

Scientific Correspondence

Scientific Correspondence is intended to provide a forum in which readers may raise points of a scientific character. Priority will be given to letters of fewer than 500 words. Manuscripts can be submitted to either London or Washington.

Nomenclature for life

SIR — The use of the name Hadean for the first aeon of the Earth's history¹ has become widespread, but the usage of the term remains variable. In recent years, minerals and rocks roughly 4,000–4,300 million years (4–4.3 Gyr) old have been found on Earth, and work on ribosomal RNA² has provided insight into the earliest history of life. With this new interest, it becomes necessary to fix the terminology so that the very diverse community of researchers studying the early Earth can use common nomenclature.

The Hadean is the time of Hades; the next aeon, the Archaean, takes its name from the Greek (as in the opening of St John's gospel: "En Archi" or "In the beginning"). The most obvious moment to fix the boundary between the Hadean aeon and the Archaean aeon is at the start of life³. The recognition of the domains of life² allows a more precise definition, however; we should place the end of the Hadean and the start of the Archaean at the time when the last common ancestor (a single cell, the common ancestor to bacteria, archaea and eukarya)², or common ancestral population, of life lived.

This definition is preferable to an arbitrary time-line, both philosophically (stratigraphic definitions should be capable of location in the rock⁴) and, more to the point, practically, as the need for the definition lies mainly in the progress of molecular palaeontology.

Intrinsically, it is conceivable that isotopic fingerprints of the earliest living communities exist. For example, photosynthesis, which came later, may have left its isotopic imprint on Archaean carbonates. Until such a fingerprint is established, or a well-calibrated molecular clock is found, our best guess for the time at which the last common ancestor lived is probably roughly 4.0 ± 0.2 Gyr ago, towards the end of massive bombardment of the planet, and before the start of isotopic fractionation by photosynthesis⁵.

E. G. Nisbet

Department of Geology,
Royal Holloway College,
University of London,
Egham, Surrey TW20 OEX, UK

1. Cloud, P. C. *Trans. Geol. Soc. S. Afr.* **79**, Annexure, 1–33 (1976).
2. Woese, C. R. et al. *Proc. natn. Acad. Sci. U.S.A.* **87**, 4576–9 (1990).
3. Nisbet, E. G. *Episodes* **14**, 327–330 (1991).
4. Hedberg, H. D. *International Stratigraphic Guide* (Wiley, New York, 1976).
5. Sleep, N. H. et al. *Nature* **342**, 139–142 (1989).

1. Corcoran, D. W. Q. *Jl exp. Psychol.* **14**, 178–182 (1962).
2. Everitt, B. J., Robbins, T. W. & Selden, N. R. W. in *The Pharmacology of Noradrenaline in the Central Nervous System* (eds Heal, D. J. & Marsden, C. A.) 349–378 (Oxford Univ. Press, 1990).
3. Aston-Jones, G., Rajkowski, J., Kubiak, P. & Alexinsky, T. *J. Neurosci.* **14**, 4467–4480 (1994).
4. Robbins, T. W. & Everitt, B. J. in *Cognitive Neurochemistry* (eds Stahl, S. M. et al.) 135–176 (Oxford Univ. Press, 1987).
5. Glue, P. & Nutt, D. *Psychopharmacology* **2**, 119–137 (1988).
6. Clifford J. M., Day, M. D. & Orvin, J. M. *Br. J. clin. Pharmacol.* **14**, 99–101 (1982).
7. Broadbent, D. E., Broadbent, M. M. P. & Jones, J. L. *Eur. J. cogn. Psychol.* **1**, 69–94 (1989).
8. Middleton, H. C., Coull, J. T., Sahakian, B. J. & Robbins, T. W. *J. Psychopharmacol.* **8**, 1–7 (1994).

Renal abnormalities in mutant mice

SIR — On endothelial surfaces throughout the body, angiotensin-converting enzyme (ACE) generates angiotensin II and other peptides which contribute to fluid and electrolyte balance and to the maintenance of blood pressure. ACE is also expressed in tissues from early embryonic stages, including epithelia, where its role is poorly defined. We have been particularly interested in potential roles of ACE in development, because pharmacological blockade in neonatal rats can cause renal abnormalities¹. Targeted mutation of the ACE gene has recently been shown to reduce blood pressure in heterozygous mice and to produce infertility in homozygous deficient male mice, accompanied by renal cortical atrophy, tubular shrinkage and inflammation². Adult mice of all genotypes appear otherwise outwardly healthy².

We used gene targeting designed to preserve expression of the testicular isoform, which is generated from an alternative promoter, to make mice lacking ACE. Homozygous mutants lack ACE in all tissues except the testes. The most notable defect is the inability of homozygous mutant mice to concentrate their urine. Whereas wild-type or heterozygous littermates can concentrate their urine to more than 4,000 mOsm l⁻¹ after 2 hours of

fluid deprivation, the ACE mutant mice cannot concentrate to more than 820 mOsm l⁻¹. These changes are similar to those seen in neonatal rats treated with ACE inhibitors¹. The ACE mutant mice are also uraemic, with mean blood urea nitrogen (BUN) of 0.52 mg ml⁻¹ at 4 months of age, as opposed to 0.15 mg ml⁻¹ in wild-type littermates.

Generation of concentrated urine occurs principally in the loops of Henle and the collecting ducts of the kidney³. These sites are clearly disorganized in the mutant mice (see figure): the anatomy of the medulla (which contains the collecting system) and the pelvis (where collecting ducts empty to the ureters) is markedly distorted. The fan-like medullary rays are absent and the triangular pelvis is shrunk and misshapen. Histological section shows the medulla to be compressed by pelvic cystic dilation and medullary cysts. These changes can explain the failure to concentrate the urine. There is no evidence of ureteral obstruction. The kidneys of heterozygous mice appear normal (data not shown). Cortical atrophy, vessel-wall hypertrophy and inflammation noted previously in ACE mutant mice² are also present in these ACE mutant mice.

The ACE mutant mice reported here seem to be more sick than the ACE

mutant mice reported previously², perhaps because of the severity of the renal lesions. Although homozygous mice are born at the predicted 25% rate to heterozygous pairs, most die by the time they are weaned at 3 weeks of age. The only obvious difference between the mutants described in ref. 2 and here is that in our mutants testicular ACE is preserved. We assume that this accounts for the retention of fertility, in contrast to those lacking both isoforms². Presumably, the difference in severity reflects modification by other genes.

ACE inhibitors have been used in anti-hypertensive therapy since the late 1970s, and can be associated with renal failure in patients with compromised renal blood flow⁴. Recent reports suggest that rats exposed to ACE inhibitors in the neonatal period may suffer renal abnormalities and water wasting, similar to those reported here in the ACE mutant mice¹. Both angiotensin II and its receptors are present in the developing kidney, and it will be of interest to determine whether the products of ACE activity, which include but are not limited to angiotensin II, act as signalling molecules during assembly of the renal tubules and collecting system.

Medullary cysts, corticotubular atrophy, interstitial inflammation, water wasting and uraemia are characteristics of the nephronophthisis-uraemic medullary cystic disease complex, a group of disorders that cause renal failure in childhood and adolescence^{5,6}. In some instances it is familial, but whether any cases reflect ACE deficiency is unknown. This mouse system may offer an opportunity to study progression of this disorder and to examine directed therapies.

Charles Carpenter, Anita A. Honkanen*, Hiroshi Mashimo, Kendrick A. Goss

Paul Huang, Mark C. Fishman†
Cardiovascular Research Center,
Massachusetts General Hospital,
149 13th Street, Charlestown,
Massachusetts 02129 and

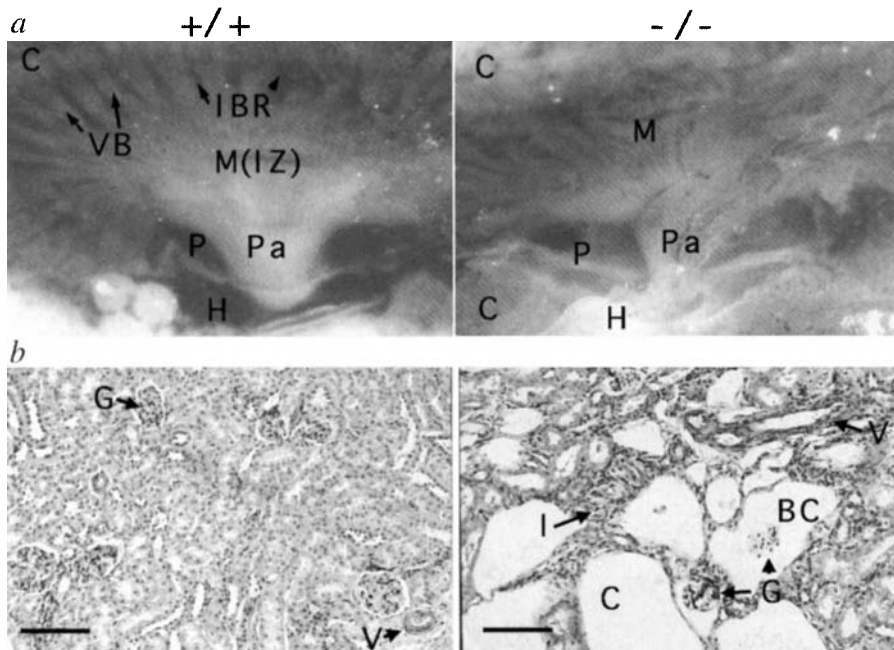
Harvard Medical School,
Department of Medicine, Boston,
Massachusetts 02115, USA

Magdi Asaad, Charles R. Dorso
Hong-sen Cheung

Department of Cardiovascular
Biochemistry,
Bristol-Myers Squibb Pharmaceutical
Research Institute, PO Box 4000,
Princeton, New Jersey 08543-4000, USA

*Also in Anesthesia Department.

†To whom correspondence should be addressed.



Cystic distortion of the renal pelvis and medulla in ACE mutant mice. *a*, Cut surface of freshly dissected kidneys; *b*, histological sections of adult kidney, stained with haematoxylin and eosin. Left columns are wild-type; right are homozygous ACE mutant mice. In the wild type, the papilla and medulla are well organized, with medullary rays in the outer medulla, whereas in the mutant there is frank disorganization of the medulla, with septate cavernous cystic spaces in the renal pelvis. In addition, the cortex has perivascular and tubulo-interstitial chronic mononuclear inflammatory infiltrates, dilated Bowman's capsules and hypertrophic vessels. C, cortex; H, hilum; IBR, interbundle region; IZ, inner zone of medulla; M, medulla; P, renal pelvis; Pa, papilla; VB, vascular bundles; BC, Bowman's capsule; C, cyst; G, glomerulus; I, inflammatory infiltrate; Scale bar, 100 µm.

À Duesberg, adieu!

John Moore

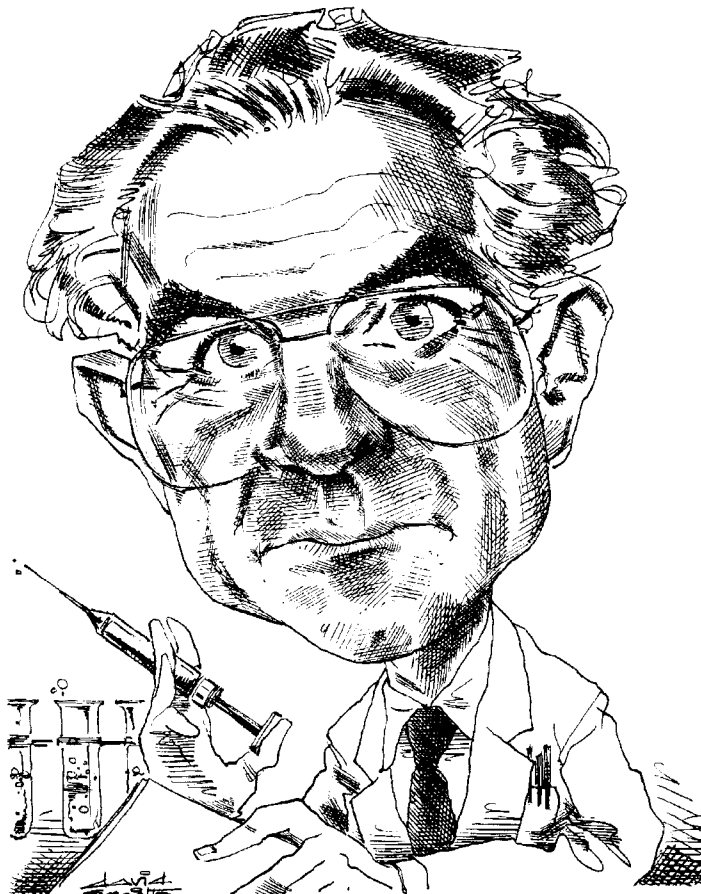
Inventing the AIDS Virus. By Peter H. Duesberg. Regnery: 1996. Pp. 722. \$24.95.

ACCORDING to Bryan Ellison, who co-wrote with Peter Duesberg an earlier version of *Inventing the AIDS Virus*, the US Central Intelligence Agency (CIA) tried to suppress the publication of this book. I can't think why it would want to bother. But conspiracy theories so pervade the book that I shouldn't be in the least surprised if Oliver Stone does the movie.

Duesberg's central thesis is that the human immunodeficiency virus (HIV) is a harmless virus, and that lifestyle (especially recreational drug use) is the principal reason why people die of AIDS. The use of AZT as an AIDS therapy is blamed for exacerbating the problem. In the first section of his book, Duesberg tells the story of an obscure syndrome (SMON) present in Japan from the 1950s to the 1970s. Despite persistent theories of a viral cause, SMON was found to be a toxicological problem caused by anti-diarrhoea drugs sometimes used to treat its symptoms. Duesberg draws an analogy between these events and AIDS, with AZT analogous to the anti-diarrhoea drugs. An interesting tale, but documenting this and a few other old medical mistakes scarcely proves that AZT causes AIDS and that HIV is a mere passenger virus. But according to Duesberg, "No fatal viral disease is known to cause death in nearly all infected people — except the paradoxical 'AIDS virus'". Try telling that to those who came across Ebola-Zaire; their mortality rate was about 80 per cent, for this virus is literally more lethal than a bullet in the head.

The book contains no new revelations on the 'non-link' between HIV and AIDS since September 1995, when the US National Institute of Allergy and Infectious Diseases released its 61-page document on *The Relationship Between the Human Immunodeficiency Virus and the Acquired Immunodeficiency Syndrome*. This contains all the facts, and I strongly recommend people to read it. Of course,

seeing that it is written by government scientists, it will no doubt be dismissed by Duesberg's sympathizers as part of a continuing cover-up. For according to Duesberg, the AIDS epidemic became the "salvation" of the Centers for Disease



Control (CDC). The Epidemic Intelligence Service (EIS) of the CDC is described as "the medical CIA" and ex-members of the EIS are said to "have obtained prominent positions in the media". One even edits a scientific journal. How sinister! Whatever next? Essentially, Duesberg's case is that the fundamental purpose of the CDC is to invent medical emergencies for the National Institutes of Health to resolve — anything is justified so long as the tax dollars just keep on rollin'. Implicit, and often explicit, is that tens of thousands of health-care professionals and research scientists are either too stupid to realize that HIV is not the cause of AIDS, or too venal to do anything about it for fear of losing income from the gov-

ernment or drug companies.

Duesberg mounts an assault on the virology "establishment", with special emphasis on the tumour virologists of the 1960s and 1970s. Researchers' mistakes, real and opined, are gleefully documented — a veritable virological *Who's Who* is castigated. And the trend continues when the HIV section is finally reached. There, all the 'big name' retrovirologists of the 1980s are targeted, and the early scandals of AIDS research are picked over yet again. So many scientists and so many of their "mistakes" are listed that I was eventually reminded of the old joke about the brigade of guards on parade, with one little guardsman horribly out of step. When the drill sergeant bawls at him, an old lady attacks him with an umbrella saying: "Leave him alone, my boy Peter is in step, it's all them other so-and-so's what are the problem!". All this ancient history is very entertaining, but it hardly seems central to the purpose of the book. Or is it?

Although some vengeance might be expected from a virologist whose eminent career was ended by the AIDS epidemic, one might have wished for a better understanding of modern virology from Duesberg. One of his main complaints about HIV and other 'slow' viruses is that they "violate the laws of virology". But what are these laws? Was it carved in stone that the Lord God spake unto the retroviridae and commanded: "Thou shalt not kill"? The great beauty of biology — indeed, of science in general — is that as knowledge advances, so paradigms shift; if HIV acts differently from the viruses Duesberg grew up with, what of it? And herein, I suspect, lies the basic problem: Duesberg clearly has an outstanding knowledge of the relatively simple avian leukaemia viruses with which he made his professional reputation. But he draws his views on how HIV 'should' behave from this early research experience; he has never published any papers based on his own work with HIV at the laboratory bench. Reading the AIDS literature can take one only so far: experimenting gives active researchers a whole new dimension to their knowledge.

I can list here only a few of the more egregious examples of Duesberg's misunderstanding of HIV virology. He states that "retroviruses do not kill cells". This assertion is not even correct for all avian leukaemia viruses, and anyone who has

cultured HIV can attest to its prominent cytopathic effects. HIV is not a leukaemia (onco)virus; it is a lentivirus, and behaves distinctly differently from the oncoviruses both *in vivo* and *in vitro*. To extrapolate from avian leukaemia virus to HIV is like asserting that because one can stroke a pussy-cat with impunity, it is perfectly safe to put one's head in a lion's mouth. Duesberg sees a fatal paradox in the fact that HIV can be grown in permanently infected, immortal T-cell lines *in vitro*, yet is supposed to cause AIDS by killing T cells *in vivo*. There is no such paradox. When a chronically infected cell culture is started, clones of cells relatively resistant to the cytopathic effects of HIV are gradually selected for and eventually take over the culture. There can also be some adaptation of the cells (and virus) to the culture conditions. The principal phenotypic change in the cells is a partial reduction in the surface expression of the HIV receptor, which reduces the extent of cell-killing in the culture. But the HIV produced in these cultures is still highly cytopathic when plated back onto unadapted primary T cells. And sadly, HIV produced from permanent cell lines is pathogenic *in vivo* — it is today causing disease in at least one

accidentally infected laboratory worker.

Duesberg writes: "Only rare luck... can extract HIV from an antibody-positive person". Perhaps I should get the technicians in our laboratory to buy my lottery tickets; they succeed in isolating HIV almost every time they try. Many of Duesberg's problems with the pathogenic effects of HIV seem to lie in his belief that HIV is dormant *in vivo*, that HIV-infected people "never have more than one in every 10,000 T-cells actively producing copies of the virus". This old canard, derived from research in the mid-1980s, has long since been proved incorrect. In the early days of HIV research, analytical techniques were obviously more primitive than they are now, so why still rely on them? The true figure for the frequency of infected cells is more like 1 in 100, although there is a wide range, depending on the state of disease progression. The documented loss of more than a hundred million T cells a day as a result of the generation of more than a billion virus particles a day attests to the virulence of HIV.

Duesberg points out that the opportunistic infections suffered by AIDS patients are unrelated to each other, and finds this hard to reconcile with any common cause, let alone HIV. The common cause is that opportunistic infections generally happen because of a dysfunctional immune system, and the cause of this dysfunction is usually HIV infection. Of course, there can be other causes — genetic or environmental — but rarely is the dysfunction as devastating as that found in the later stages of HIV infection, and never is it as common.

Duesberg believes that HIV is essentially not a sexually transmitted virus; indeed, the very cover of his book states that "AIDS is not sexually transmitted". Instead, he argues that "HIV has been passed along from mother to child for many centuries". The first statement ignores the entire body of data on the epidemiology of HIV spread in the United States and Europe, whereas the second ignores the death rate among children infected by HIV from their mothers; only a tragically small proportion of these children survive long enough to have the chance of having children of their own. How could transmission from mother to child permit sustained HIV spread under these conditions?

Much space is devoted to the thesis that AZT causes AIDS. AZT is decried as a toxic chemical, which of course it is to an extent. So are most chemotherapeutic agents used against cancer. So are paracetamol, rock salt and water if consumed in the wrong quantities. Like all drugs, AZT has a therapeutic window — a dosage that has maximum effect on its target (HIV) and minimal effect on the working of the human body. This fundamental pharmacological principle is criti-

cal for understanding AZT's (admittedly limited) effect on HIV replication *in vivo*. Adding human interest to an otherwise dry section are the numerous quotations from people who believe that AZT has harmed them or their infants. But what of Elizabeth Glaser, who later founded the Pediatric AIDS Foundation? She was infected by HIV through a blood transfusion, and then passed the virus to her children. None of the family used recreational drugs. Sadly, Elizabeth and her daughter Ariel eventually died of AIDS. But at a critical stage of Ariel's disease, Elizabeth managed to obtain AZT for her unconscious child. I quote below from *In the Absence of Angels*, Elizabeth's book: "Three weeks to the day after we started intravenous AZT I walked into Ariel's room in the morning and she looked up and said 'Good morning, Mom. I love you'.... She hadn't talked in three months!.... It was the miracle we had been waiting for". No AIDS researcher pretends that AZT is the answer to AIDS. But neither is it the cause of it. Most people who die of AIDS have never taken AZT or any other Western drugs. Neither have the monkeys that die from AIDS induced by molecular clones of SIV, a lethal close cousin of HIV.

Duesberg wraps together his twisted facts and illogical lines of argument to create a tangled web to trap the unwary, desperate or gullible. But however much he attempts to gild his writings with philosophies of scientific truth, the reality is that his premises are based not on facts but on faith: faith that he is right, and everyone else is wrong. This was his position long before AIDS appeared, as tumour virologists know well.

Duesberg ends by detailing his ostracism by the virology community, his inability to get research funding, the personal snubs he has suffered. The advent of HIV has clearly been a personal tragedy for a once highly respected retrovirologist, but one's sympathy must of course be tempered by thoughts of those for whom AIDS has been a rather greater personal tragedy. Three years ago, I likened Duesberg to the Black Knight from "Monty Python and the Holy Grail". This character had his limbs hacked off one by one, but the game little torso tried to bite the knee-caps from his assailant. The events of the past few years have extracted the Black Knight's teeth, leaving him with the sole recourse of spitting at those whose views of virology have differed from his over the past two decades. But where the spittle lands is on the graves of those millions of people killed by HIV, and on those it has yet to slaughter. How sad, and how ultimately pathetic. □

John Moore is at the Aaron Diamond AIDS Research Center, 445 1st Avenue, New York, New York 10016, USA.

Also on AIDS

AIDS: Virus- or Drug-Induced? edited by Peter H. Duesberg, Kluwer, DF175, \$49.50, \$33.75. Two dozen scientists, scholars and journalists question what they see as the "status quo" of AIDS research. Reprints selected correspondence between *Nature* and Duesberg.

AIDS in the UK: The Making of Policy, 1981-1994 by Virginia Berridge. Oxford University Press, £45 (hbk), £12.99 (pbk). A scholarly and immensely detailed account of the reaction of the political and medical establishments, and of the public, to the advent of AIDS in the United Kingdom.

The Gravest Show on Earth: America in the Age of AIDS by Elinor Burkett. Houghton Mifflin, \$22.95. A journalist's account "of science run amok, of backroom deals between activists and government bureaucrats, of biotechnology companies manipulating stock prices by manipulating research results. It is a tragedy of bungled research, scientific and activist vanity, the dangers of political correctness and, finally, the price more than a million Americans are paying for a nation's folly."

WHO Model Prescribing Information: Drugs Used in Sexually Transmitted Diseases and HIV Infection. World Health Organization, Sfr25 (developing countries, Sfr17.50) (pbk). A practical guide for clinicians to the selection and prescribing of essential drugs, especially in developing countries.

Money matters

David Pearce

The State of Humanity. Edited by Julian L. Simon. Blackwell: 1996. Pp. 694. £55, \$54.95 (hbk); £16.99, \$22.95 (pbk).

From Care to Action. By Martin Holdgate. Earthscan: 1996. Pp. 346. £35, \$56 (hbk); £14.95, \$26 (pbk). Distributed in the United States by Island.

As the millennium approaches, it is time for reflection and anticipation. Some might argue that humankind has made such a mess of things that it cannot seriously expect to survive for much longer, at least in current social forms. Only those who believe in some evolutionary goal for humans may be upset by such a prospect: after all, the end of humankind will not be the end of all species, although many may go down in the process. Earth surely transcends the human race. Others are less doleful and seem divided into two camps: those who recognize there is a serious problem, but who believe there is a chance of a sustainable and reasonably tolerable future if only we act positively and now; and those who deny there is a problem, that scaremongers have exaggerated the issues and that the many social and economic feedback mechanisms will correct, sooner or later, any problems that do arise. Call them the concerned and the benign optimists respectively. Martin Holdgate belongs to the former group, Julian Simon to the latter. These two volumes could hardly be more different.

Simon's book has nearly 60 contributions from 64 authors, 62 of them from the United States. Around half the contributions are on environmental and natural resource problems, but there are essays on, among other things, smoking, murder, inequality in the United States, poverty and homelessness. The US bias shows. For example, Bruce Gardner and Theodore Schultz's essay on soil erosion relates to the United States, not developing countries, where many would argue the issue is far more serious. Laurence Kulp's essay on acid rain declares that it is "clearly beneficial" to crops in the United States, and has only "small negative effects" on buildings and materials, neither of which conclusion is consistent with European studies. Occasional forays into non-US evidence also reveal the risks of a dominant US authorship: thus William Baumol and William Oates, both immensely distinguished economists, rightly applaud the restored water quality of the River Thames in the south of England, but might be surprised to learn that, at least in terms of length of rivers classified as clean, there has not been a very discernible improvement over the past 30 years in the other rivers of England and Wales.



In February 1994 the last remaining huskies were banished from Antarctica, forced into retirement by the introduction of mechanical snow vehicles and pressure from environmentalists. The story of the dogs that served the British Antarctic Survey for over 50 years is now told in the richly illustrated *Of Dogs and Men* by Kevin Walton and Rick Atkinson. Images, £19.95.

Those who know Simon's previous work will not be surprised at his upbeat introduction. It is worth quoting:

Almost every absolute change, and the absolute component of almost every economic and social change or trend, points in a positive direction, as long as we view the matter over a reasonably long period of time. That is, all aspects of material human welfare are improving in the aggregate.

And he offers to bet a week's or a month's pay that these improving trends will continue indefinitely. Many indicators support his contention that there have been unquestioned improvements in the past, but it may be an unwise man who bets on the future. As Simon shows, life expectancy has increased enormously, the real price of resources — one indicator of scarcity — has fallen, literacy has improved, nutrition has improved, pollution has declined. Some of the indicators are global but some are peculiarly parochial. There is no mention, for example, of the fact that the average increase in per capita gross national product was close to zero for the 40 or so lowest income countries in the world between 1980 and 1993, excluding China and India. Perhaps this is too short a period to detect

a long-run trend. Where the numbers do not support the case, the numbers are questioned. Concern about species extinction, for example, is an "environmentalist scam". Simon, writing an essay with the late Aaron Wildavsky, may be right to question the generality of 'biogeographical' or 'species-area' theories of extinction whereby extinction is inferred, not observed, from data on land-use conversions. And he is right to say that the actual recorded rate of extinction is around one species a year since 1600, a result confirmed in the United Nations Environmental Programme's recent *Global Biodiversity Assessment*. This hardly seems dramatic. But more to the point is the number of threatened species today — more than 5,000 animal species and some 26,000 plant species — and the rate at which known species go from being unthreatened to threatened. This is the view adopted by Holdgate, and on this basis it is hard to be as sanguine as Simon.

There are good features of *The State of Humanity*. First, it contains a mass of information and some of it will be unfamiliar. Second, there are some thoughtful and fascinating essays, for example on an analysis of long-term trends in suicides, on long-term trends in standards of living and

on inequality in the United States (although this neglects the startling reversal in the past few decades). Third, it rightly takes to task the self-appointed doomsters who perhaps do not deliberately exaggerate, but who exaggerate because of poor research and inadequate questioning. The general *laissez-faire* tone of the volume will offend some, but, if so, the offended should be prepared to answer the case made in the book.

Holdgate accepts there is an array of environmental problems and spends more time telling us what has to be done. The recommendations are familiar but no less important for that: empowering communities, moving from outright conservation to sustainable use, education and new alliances between interest groups. He is sincere in his belief that we need a 'new ethic' of care, but the menu of action looks strangely naive: building a new conscience and spirituality, reforming the international economic order, and "new economic thinking" (when we have hardly made use of the power of the 'old' economic thinking) to improve our lot and the lot of other species.

Holdgate's volume is a worthwhile read but it has an inner risk. Spiritual shopping lists calling for 'changes in values' invite others to philosophize and debate while the environment burns. Holdgate is wise enough to see some of this risk and he calls for other action too. But one wonders if the dreamers really understand that it is all about competition between humans and other species for space, for a slice of the Earth. Until environmental assets secure economic value greater than that embodied in the competing uses of the land on which they rely, the state of humanity really is at risk, as much, it seems, from the concerned as from the benign optimists. □

David Pearce is at the Centre for Social and Economic Research on the Global Environment, University College London, Gower Street, London WC1E 6BT, UK.

New in paperback

The Strange Case of the Spotted Mice and Other Classic Essays on Science by Peter Medawar. OUP, £7.99.

From Here to Infinity: A Guide to Today's Mathematics by Ian Stewart. OUP, £7.99. A revised and retitled edition of *The Problems of Mathematics*.

Natural Acts: A Sidelong View of Science and Nature by David Quammen. Avon, \$11. "Quammen's book is full of strange information about animals and their habitat, about Tycho Brahe's artificial nose, and other such oddities. His exuberance carries one, sometimes unwillingly, along", wrote Walter Gratzer in a review in *Nature* **318**, 116 (1985).

Reality bites

Freeman J. Dyson

The End of the World: The Science and Ethics of Human Extinction. By John Leslie. Routledge: 1996. Pp. 305. £16.99, \$23. To be published in the United States in May.

THE subject of this book is a philosophical speculation that the author calls the "doomsday argument". The same speculation was independently put forward in *Nature* (**363**, 315; 1993) by Richard Gott, who gave it the more appropriate name of the Copernican principle. The idea is to estimate the probability of the survival of the human species by using the Bayes rule. The Bayes rule says that (probability of A) multiplied by (probability of B given A) is equal to (probability of B) multiplied by (probability of A given B). The rule is supposed to apply to any two hypotheses A and B, because both sides of the equation are equal to the probability of the hypothesis (A and B). Formally, the rule is almost a tautology. But when it is applied to hypotheses in the real world, it can easily lead to absurd conclusions, because of ambiguities in the meaning of probability. Every estimate of probability must be based on a certain division of the world into known and unknown. The Bayes rule is valid only if the four probabilities in the equation are all based on the same division of the world. The phrase (probability of A) has no absolute meaning independent of previous knowledge.

Leslie applies the Bayes rule to a variety of situations of which the following is typical. Hypothesis A says that the human species will ultimately give birth to less than X people. Hypothesis B says that Leslie and his readers live at a time when the human species has given birth to fewer than Y people. Evidence from history and palaeontology indicates that (probability of B) is close to unity when Y is of the order of a hundred billion. Also, (probability of A) cannot exceed unity. Thus the rule implies that (probability of A given B) does not exceed (probability of B given A). Now, the probability that Leslie and his generation should happen to be among the first Y people in order of birth, out of a total population greater than X , is less than Y/X . That is to say, (probability of B given A) is less than Y/X . Therefore the Bayes rule leads to the conclusion that (probability of A given B) is also less than Y/X . That is to say, there is less than a ten-per-cent chance that the human species will survive long enough to give birth to a trillion people. The conclusion is that our species has only a ten-per-cent chance of surviving as long as five hundred genera-

tions with our present size of population, or fewer than five hundred generations if the population grows larger. Leslie elaborates this conclusion with numerous variations. He examines many different hypotheses and deduces the corresponding probabilities of survival. But the essential core of his reasoning is as I have stated it. The reasoning stands or falls according to whether the application of the Bayes rule is valid or invalid.

After careful consideration, I state unequivocally that the application of the Bayes rule is here invalid. The reason the rule fails is simple. The two quantities (probability of B) and (probability of B given A) are not based on the same division of the world into known and unknown. When we estimate (probability of B) to be close to unity, we assume that we know enough about palaeontology to be sure that we are living among the first hundred billion humans. When we estimate (probability of B given A) to be small, we assume that we know nothing of our place in the history of our species. The two estimates of probability are based on incompatible assumptions and are therefore not comparable. The application of the Bayes rule is invalid because the rules of the probability game have been changed between the left and right sides of the equation.

Half of the book is devoted to theoretical discussion of the Bayes rule and its application to the question of human survival. This discussion is worthless, being based on an invalid use of the rule. The other half of the book is a recital of the many real dangers to which our species is exposed. Here Leslie has nothing new to add. He refers to a large number of sources from the literature of science and risk assessment, but does not examine them critically.

Almost two hundred years ago, in 1798, Robert Malthus published his famous *Essay on the Principle of Population as It Affects the Future Improvement of Society*, deducing dire predictions of future misery from two hypotheses that he called the geometrical increase of population and the arithmetical increase of subsistence. His predictions of inevitable poverty failed because his hypotheses were faulty. The increase of subsistence turned out to be much faster than arithmetical. Nevertheless, uncritical belief in Malthus's predictions helped to hold back political and social progress in Britain for a century. Because of this unhappy precedent, I consider it important to call attention to the fallacy in Leslie's argument. A mistaken philosophical argument may have serious consequences in the real world. □

Freeman J. Dyson is at the Institute for Advanced Study, Princeton, New Jersey 08540, USA.

Limits of safety

Harry Collins

The Challenger Launch Decision: Risky Technology, Culture, and Deviance at NASA.

By Diana Vaughan. University of Chicago Press: 1996. Pp. 575. \$24.95.

No Downlink: A Dramatic Narrative About the Challenger Accident and Our Time. By Claus Jensen. Farrar, Straus, Giroux: 1996. Pp. 397. \$25.

THE space shuttle Challenger blew up on 28 January 1986, when exhaust gases escaped through a joint in one of its solid-propellant rocket boosters, burnt a large hole in the side and ruptured the fuel tank holding liquid hydrogen and oxygen. It is now thought that cold weather caused the rubber O-rings, which should have sealed the joint, to lose their resilience. Three images seem to tell the story: the diverging pattern of smoke in the sky; Richard Feynman, America's favourite physicist, dipping O-ring material in a glass of ice-water and showing that it stiffens with cold; and anxious managers, struggling to maintain the public credibility of the ailing NASA (National Aeronautics and Space Administration), overruling the engineers who wanted to veto the launch because of the cold. With this accounting, the tragedy became a triumph for technology and homespun genius; all the failure can be found in the politics. Like a folk tale, the story supports a mythical version of scientific and technological practice — a myth that has outlived its usefulness.

Claus Jensen is a Danish professor of literature and his writing holds one's attention throughout. His opening chapter covers the launch and termination of the fatal Challenger mission. The phrase 'no downlink' refers to the severing of radio contact after the explosion as well as the severing of management from its engineer subordinates. He sets Challenger into the context of Charles Perrow's idea of 'normal accidents' — accidents that are inevitable in any complex technological organization. Some revelations about Challenger take one's breath away. For instance, there was the blatant interference of a senator from Utah in giving the contract to home-based Morton Thiokol in the face of bids from contractors with more suitable experience, an example of what Americans call 'pork-barreling'.

But other 'obvious' mistakes were all too normal in a system as complex as this. Jensen's book is also full of the sort of astonishing facts that bring out the school-child in all but determined cynics: the shuttle's turbo-pumps can empty an Olympic swimming pool in 25 seconds; its insulating tiles can be balanced comfortably in the hand even when one side is

red-hot; and its solid rocket boosters burn around ten tons of fuel a second. One of the reasons the book is so gripping is that the professor of literature cannot resist a beginning, a middle and an end. Feynman is the hero who reveals the management's

IMAGE
UNAMAGABLE
FOR A COMPLAINT
FOR REASON
REASONS

Challenger's lift-off on its ill-fated mission. A puff of black smoke can be seen at the base of the solid rocket booster where an O-ring seal failed on ignition.

mistakes, which were made in the climate of America's undiminished technological ambition and pursued within a diminished financial regime. The pressure on NASA was translated into pressure on subcontractors, so that when the engineers warned that the solid booster seals could not be guaranteed at the unprecedentedly low temperatures of the launch, they were overridden by Thiokol decision-makers scared of NASA. No-one could face another costly, exhausting and embarrassing postponement. It all seems quite tidy.

Diane Vaughan, a professor of sociology, is more ambitious. Where Jensen worked from written sources available in Denmark, she has delved deep into the archives and conducted interviews with

many of the protagonists. Typically American, her book is long and stuffed full of technical detail, engineering charts and diagrams. Her unapologetic presentation of the technicalities is the book's main strength. It makes for a less exciting read than Jensen's but, to use the anthropological jargon, it is the 'thick description' that demythologizes the retro-spective account of the Challenger tragedy. Jensen tells a fascinating story, but Vaughan advances an important thesis.

Vaughan juxtaposes the backward-looking account with a version of events embedded in the rich and complex context of day-to-day engineering life. She takes 55 pages to cover the eve-of-launch discussion. We learn that it is not simply that financial and administrative pressure forced engineers and managers to ignore their better judgement. We learn that it is not just that anyone would have noticed the O-rings would be stiff when cold if only they had done something as simple as dip them into ice-water — as in Feynman's limelight-grabbing performance on television. Everyone knew the O-rings stiffened as they cooled, but they did not know how serious the problem was. The charts of stiffness discussed before the launch showed that the Viton rubber did not stiffen catastrophically until below 25 degrees Fahrenheit. Admittedly, a group of engineers, the most well known being Roger Boisjoly, recommended that there should be no launch below 53 degrees, but they did so on the basis of having experienced only two other cases of O-ring erosion, one of which took place at 75 degrees.

So, before the launch decision, erosion had been experienced during both the coldest and the hottest launch, and that is why the managers, whose job it was to cope with the conservatism of engineers, might reasonably have concluded that unprecedented cold was not necessarily going to be fatal. In any case, during the build-up to the preceding launch, temperatures well below 53 degrees had been experienced during countdowns and not a word of warning had come from the Thiokol engineers. Further, in the immediately preceding Challenger countdown, called off because of high cross-winds, the temperature had been 38 degrees, yet the cold had not been discussed. Again, not all the Thiokol engineers agreed, but it is only the testimony of those who got it 'right' that has been given salience. So the sudden attempt to impose a cold-weather veto was bound to seem infuriating and arbitrary to the managers, especially as the engineers did not press home their case against the counter-view.

In retrospect, of course, the decision was wrong. In retrospect we can see that

NASA/SPL

the O-ring joint should have been redesigned years before. The O-rings were known not to perform according to the original design because the thin casing flexed unexpectedly and the joints tended to widen when the rockets exploded into life. But the O-rings had always worked, and the pragmatic approach became 'normalized'. In retrospect it was not so difficult for Feynman to show that the shuttle was a far more dangerous machine than the American public had been led to believe. But in retrospect it was a far more dangerous machine than anybody believed, or they would not have launched it. It was nobody's intention to kill the astronauts or lose their own jobs.

Feynman concluded that, given the state of the shuttle at the time of Challenger, the chances of disaster from one cause or another were between one and two in a hundred — that is, the shuttle programme could expect between 1 and 2 per cent fatalities. Subsequently, NASA has spent a great deal of money to reduce the risk, but I wonder if it is worth it.

I have been lucky enough to watch a shuttle launch (the night-launched Hubble Space Telescope repair mission). I found myself deeply moved by fearful empathy for the fragile humans sitting on that column of fire gradually dwindling to

a distant star, but I would volunteer for the ride in a trice even after reading these two books. Thirty workers were killed building the Golden Gate Bridge, and they were not volunteers. If only we could find a way of spending the shuttle improvement money on those who take risks because that is the only way they can earn a living, accept that it is in the nature of pioneering technology to be risky and get on with it. Of course, when it is not volunteers' lives that are at risk but communities of innocent people, as in nuclear power engineering, the moral is completely turned around; in such circumstances those who argue that complex technology can be made safe are deeply culpable.

The danger of the simple folk tale of Challenger, which safeguards the science at the expense of the managers, is that it strengthens the idea that complex technology is perfectible. Vaughan's book shows why we should not believe it; we make our technology and we can control and understand it, but only to the extent that we can control and understand ourselves. □

Harry Collins is at the Centre for the Study of Knowledge, Expertise and Science, University of Southampton, Southampton SO17 1BJ, UK.

Self-fulfilling prophecies

David Edgerton

Knowing Machines: Essays on Technical Change. By Donald MacKenzie. MIT Press: 1996. Pp. 338. \$35, £22.50.

ENGINEERS have long argued that one of the principal problems of modern society has been its slow adaptation to the needs of the new technologies they created. In the nineteenth and twentieth centuries, some engineers such as Herbert Spencer went on to be pioneers of the new science of sociology. After 1945, engineers urged the development of industrial sociology and the taking into account of the 'human factor' in industry. They generally took the view that modern technologies appeared almost out of nowhere, and at an advancing rate, but societies and cultures had no tendency to change, except belatedly under the influence of new technologies. The resulting 'cultural lag' meant not only that technologies were too slowly applied but, more importantly, that they were inappropriately applied: to war, to the creation of unemployment and so on. The scientific study of society promised to produce a kind of social engineering that would reduce the lag and ensure that technology was properly used.

The same sort of arguments are still to be found in popular discussions of tech-

nology and society; we are constantly told we are not ready for this or that technology and that public understanding is deficient. But academic sociology's approach to technology has moved on. In the late 1960s and 1970s, notably in Edinburgh, a group of sociologists, historians and philosophers argued that sociology, born out of science, should be applied to the study of science itself. This was a daring and controversial move, not least because it suggested that social factors went into the making of scientific knowledge. The 'Edinburgh School' developed — from the standard canons of scientific method — a series of prescriptions as to how one should study science. Perhaps the most famous was the symmetry principle: that the same analytical tools should be applied to studying all science, whether 'good' or 'bad'.

In the case of technology, this prescription amounts to saying that we should not explain the success of a technology by assuming its intrinsic superiority, while accounting for the failure of good technology by invoking social factors. Both technical and social factors should be involved in discussions of success and failure. Donald MacKenzie has been one of the pioneers of applying these powerful tools to technological knowledge (both formal

and tacit) and to technological innovation in particular. The approach yields a rich account of invention and innovation far from the banal stories of heroic invention by unique people 'ahead of their time'.

MacKenzie's introductory essay provides the clearest guide I know to the methods and concerns of the new sociology of technological knowledge and innovation. It is especially attractive because it deals seriously with a least some of the criticisms of his approach, and distances itself from some of the excesses. MacKenzie focuses on such issues as who new technology is best for, the path-dependence of innovation, the need for symmetrical explanation of success and failure, the importance of self-fulfilling and self-negating prophecies in technological innovation and the importance of the study of formal and tacit knowledge in technology.

Reprinted from a wide variety of sources, including *Nature*, *Technology and Culture* and the *American Journal of Sociology*, the essays in the book cover an amazing mix of topics in great depth and with great subtlety. MacKenzie is a fluent writer who avoids jargon and presents theoretical arguments lucidly. His historical narratives are tightly constructed and rich in detail.

Many of the studies are concerned with the history of computation, especially in relation to the military. The author outlines the influence of two laboratories, Los Alamos and Livermore, on the design and specification of supercomputers. They defined, he argues, what a supercomputer should be. He also provides an acute biographical sketch of Seymour Cray (written with Boelie Elzen), the most famous supercomputing entrepreneur.

Sociologists of knowledge have been traditionally very alert to controversies in science and technology, because they provide a glimpse into the nature of science and technology that is closed off when sciences and technologies settle down to a consensus. Of particular interest are MacKenzie's accounts of debates about computer-generated proofs of the integrity of computer programs.

MacKenzie is also concerned with competition between technologies, and provides a fine account of the development of the laser gyroscope for inertial navigators in civil aircraft. His point is that the early laser gyros were not manifestly superior or even superior at all to mechanical gyros. But it was believed that they would be and as a result they partially became so, because of the shift of development resources from mechanical to optical systems. So we should not assume from the ultimate success of the laser gyroscope that its victory was a foregone conclusion. Things could have turned out differently.

One of the great propaganda clichés of the nuclear age is that 'we can't uninvent nuclear weapons'. MacKenzie (writing in

this case with Graham Spinardi) shows that this argument rests on a very particular account of the nature of science and technology. They argue that the science and technology of nuclear weapons relies on the constant performance of particular sorts of experiments, and a great deal of highly specific tacit knowledge. Nuclear technology has been built up in only a very few states, and with extraordinary difficulty. The capacity to build weapons could in practice disappear, even if all the science stayed in the public domain. The ultimate science-based weapon relies on much more than written knowledge and standard engineering capabilities.

In some ways MacKenzie is attacking

straw men in his discussion of technological innovation, but this is not really a criticism of MacKenzie. It is, rather, a criticism of the way that we think and write about technology. We look at technologies in much the same way as they appear in advertisements: as fully formed, perfect and inevitable developments following a logic of technical development. The real world of new machines is very different, as we know from our own experience as well as from historical studies and MacKenzie's detailed empirical work. □

David Edgerton is at the Centre for the History of Science, Technology and Medicine, Imperial College, London SW7 2AZ, UK.

Taking things at interface value

Graham Farmelo

Software for the Self: Culture and Technology. By Anthony Smith. *Faber: 1996. Pp. 128. £7.99, (pbk).*

Cultural Babbage: Technology, Time and Invention. Edited by Francis Spufford and Jenny Uglow. *Faber: 1996. Pp. 290. £15.99, \$24.95 (pbk).*

ACCORDING to Oscar Wilde, "all art is quite useless", and, to the extent that such a brutally utilitarian view is valid, the same is true of culture. Because part of the value of culture lies in its transcendence of the material world, it appears to be very different from technology, which is concerned only with the materially useful. What, then, is the place of technology in culture?

Information technology — the defining technology of our age — poses particularly difficult problems for the cultural theorist. Something so widely available to so many is not readily accommodated in a view of a culture based on what that austere critic Matthew Arnold identified as "the best that has been known and said in the world". It is unlikely that he would have found much "sweetness and light" in the hegemony of the silicon chip.

In *Software for the Self*, Anthony Smith examines the slippery concept of culture and examines its meaning in the light of modern technological developments in the media. The book is a collection of thoughtful and eclectic essays based on lectures the author gave in 1994, in a series named after T. S. Eliot, Arnold's twentieth-century standard-bearer. Smith begins by explaining how the 'arts' — that broadest of portmanteau terms — came to be central to the usual narrow idea of what constitutes cultural activity. This leads him to analyse how the century's most pervasive cultural revolution has been effected electromagnetically, through the introduction of broadcasting media.

Smith makes a powerful case for the

importance of the concept of information in analysing the modern cultural agenda. He points out that information is a classic example of a transforming concept that, like chaos, the unconscious mind and the uncertainty principle, "rapidly invades every science, and its analogues and derivative images spread through the arts and into political discourse, eventually arriving at the gates of government". Scarcely a branch of modern industrial society has not been transformed by its own contact with the new paradigm of information, and we are all trying to come to terms with the potential of a global communica-

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Prototype computer: Babbage's first "Difference Engine".

tions network, which promises to make most media readily available to anyone with a computer.

Smith points out that, with virtual reality still in its infancy and the potential for interactivity with the conventional media still largely unexplored, there are more cultural upheavals in the offing. Yet he pays little attention to the effects that the growth in the quantity of information in the media have had on the quality of the material purveyed. Has anything of significant literary worth ever been created in cyberspace? Perhaps it is a coincidence

that it is in the adolescence of the digital age that such a question is regarded as misconceived or, at least, *outré*.

Smith has given us a valuable new perspective on contemporary culture in these fine essays, all happily devoid of jargon and snobbery. It is regrettable that he been so badly let down by his publishers. They tell us next to nothing about him, fail to provide him with an index (essential in a work of this breadth) and make no effort to package his book so that it might stand a chance of attracting the wide audience it deserves.

The unity of purpose that gives such collective power to *Software for the Self* is not evident in *Cultural Babbage*, a collection of pieces that idiosyncratically explore connections between science and technology (rarely distinguished here, regrettably) and various branches of the arts and humanities. Its 14 contributors include critics, poets, biographers and historians, each of them plainly at home in both of C. P. Snow's 'two cultures', in so far as they exist separately.

The theme of most of these essays is the place in the firmament of culture occupied by technology, in particular by machines and other inventions. As the title of the collection hints, a central figure here is Charles Babbage, the irascible polymath best remembered for his failure to build a complete version of what would probably have been regarded as the first computer. Doron Swade's elegant and scrupulously fair analysis of Babbage's reputation is followed by the highlight of the volume,

Simon Schaffer's "Babbage's dancer and the impresarios of mechanism". This densely argued piece looks at how the tough-minded Babbage and his peers were fascinated by apparently frivolous automata, such as the silver danseuse that he exhibited, close to the unfinished portion of his first 'Difference Engine', in his London salon. Schaffer links this fascination with the perception of machine intelligence, which he convincingly argues "hinges on the cultural invisibility of the human skills which accompany them".

The quality of the other pieces is uneven but they do include some gems, notably Lavinia Greenlaw's look at the connections between science and poetry, Neil Belton's analysis of the British scorn of the scientific, and Jon Katz's advocacy of the republican revolutionary Thomas Paine as the spiritual father of the Internet. The excellent index makes the collection a valuable work of reference, although the bibliographical information is disappointing, to say the least.

In her introduction, Jenny Uglow remarks that the essays are "ballasted by

detail but untethered by the ropes of academic apparatus" so that "they let the balloon of ideas float free across the orthodox science/arts divide". During our repeated crossings of the Snow-line, we enjoy a stimulating and intermittently exhilarating ride, although it is never quite clear where the balloon is heading. □

Graham Farmelo is at the Science Museum, Exhibition Road, London SW7 2DD, UK.

Body work

David Goodstein and Judith Goodstein

Twentieth Century Physics: Volumes I, II and III. Edited by Laurie M. Brown, Abraham Pais and Brian Pippard. *Institute of Physics/American Institute of Physics: 1995. Pp. 2,059. £250, \$400 (3 vols).*

THIS book is a monumental undertaking. In fact, speaking of undertaking, it is an obituary written before the body is cold. A few even think the heart may still be beating. Nevertheless, it will serve a number of important purposes.

With more than 2,000 pages in three volumes, *Twentieth Century Physics* is comprehensive. It opens with an essay by Brian Pippard on "Physics in 1900", followed by Abraham Pais on "Introducing Atoms and their Nuclei" and, a bit later in the volume, Laurie Brown on "Nuclear Forces, Mesons, and Isospin Symmetry". The third volume closes with philosophical essays on the history, substance and social structure of twentieth-century physics by Philip Anderson, Steven Weinberg and John Ziman. In between, the various fields of physics and closely related sciences (medical physics, geophysics) are covered by their practitioners. The level is uncompromisingly technical. This is not a reference for schoolchildren or journalists.

What is it then? According to the editors, they have created it because the history of physics is too important to be left to the historians. This, they imply, is history as it ought to be written, done by the pros, the participants who are qualified to decide what matters and what doesn't. In short, this is the peculiar view of history seen in a thousand physics textbooks, distilled to an undiluted essence. If nothing else, it shows us vividly how physicists wish to see their craft portrayed.

Physics, we learn, does not depend on the people who discovered it, or on the conditions in which they lived. Pais quotes a biographer of Rutherford: "People ask... what was the man like, did he really look like a farmer.... None of these things really matter." Physics, then, in a first approximation, is a body of knowledge about things,

thoroughly drained of flesh and blood.

Ah, but that won't do of course. The physicists' egos won't permit it. Impersonal phenomena, such as Compton scattering or the Mössbauer effect, get named after real people. In these volumes, the technique for reintroducing life onto the barren planet of physics is some 63 biographical sketches, set off in boxes, scattered through the three volumes. In fact, this is one of the most important functions that the editors and their contributors have taken on. This is an anointing of the winners, and an omitting of the losers. It is, at the very least, an updating of the scoreboard in the great and noble competition of physics.

Of the 63 sketches, 61 are about white, European or American males (Ryogo Kubo and Hideki Yukawa are the exceptions). Incredibly, Robert Millikan is omitted, although his student Carl Anderson is included (can the positron really be more important than the electron?). Beyond this galaxy of superstars, every physicist who picks up this set of volumes will flip first to the name index to see if he's made it (they are nearly all he's). Brown, Pais and Pippard had better be prepared to take some heat.

It is hard to imagine a history of anything in the twentieth century that omits completely to mention that great cataclysmic event at the middle of the century, the Second World War. This is especially true of physics, where the war had immense effects, both positive and negative. *Twentieth Century Physics* tries hard to ignore it, and almost succeeds, but not quite. For example, of Max Planck, we are told that "he defended German science in the difficult times of the Third Reich", and we learn that Nernst's school, "sadly depleted by Nazi anti-Jewish measures, flourished in exile". Nevertheless, the Second World War doesn't make the subject index, and neither does Los Alamos.

All of the above is more or less made inevitable by the nature of the subject and the task that the editors and contributors set for themselves. Within the limits of that task, their accomplishment is impressive. As one might expect, the contributed articles are a bit uneven. Most read like physics review articles, typically found in the journal *Reviews of Modern Physics*. All are professional and authoritative, although a few tend to re-draw the map slightly to place the author closer to the centre. Dividing up the subject matter must have been difficult. For example, the subdiscipline that calls itself 'condensed matter physics' is represented by articles on crystallography by William Cochran, statistical mechanics (largely critical phenomena) by Cyril Domb, nonequilibrium statistical mechanics by Max Dresden, superfluids by A. J. Leggett, phonons and spin waves by R. A. Cowley and Pippard, magnetism by K. W. H. Stevens, electrons

in solids by Pippard, materials by Robert Cahn and polymers and the like by P. G. de Gennes. Cross references are achieved by boxes in the margin saying "see also page xxx".

Nobody is going to read through these three volumes from cover to cover. Not even we managed that, despite heroic efforts and gallons of black coffee. But many physicists will dip into them and read sections with real pleasure. The work will also provide source material to spruce up courses and seminars with a bit of historical introduction. All this is very much to the good. The only remaining question is will it serve the main purpose the editors had in mind?

It seems clear that the editors intend their volumes to survey the territory and lay out the road maps and instructions by means of which future historians of science will find their way around physics in the twentieth century. Given the generally abysmal state of the secondary historical literature in this field, they may well succeed, up to a point. These essays are not a bad way to get an initial overall orientation to what happened in a given area. But real historians will want much more than is available here. Physicists can and do speak with confidence and authority (some would say arrogance) about their discipline, and this series is a perfect example of that. Nevertheless, this much will remain true: historians always have the last word. When historians write the story of twentieth-century physics, they will put the flesh and blood back into it. □

David Goodstein and Judith Goodstein are at the California Institute of Technology, Pasadena, California 91125, USA.

New and notable

Fire in the Mind: Science, Faith and the Search for Order by George Johnson. Viking, £18. A *New York Times* science writer's overview of some of the emerging new ideas in physics and biology coming out of Los Alamos National Laboratories and the Santa Fe Institute in New Mexico.

Charles Darwin's Letters: A Selection 1825-1859 edited by Frederick Burkhardt. Cambridge University Press, £45, \$54.95 (hbk), £15.95, \$19.95 (pbk). Letters taken from the monumental Cambridge edition of *The Correspondence* which reveal the human story behind the *Origin of Species*.

Dinosaur in a Haystack: Reflections in Natural History by Stephen Jay Gould. Harmony/Cape, \$25. The author's seventh collection of masterly musings on evolution and other natural phenomena, culled from his monthly column in *Natural History* magazine.

Testimony to niceness

Andrew Whiten

Good Natured: The Origins of Right and Wrong in Humans and Other Animals. By Frans de Waal. Harvard University Press: 1996. Pp. 296. \$24.94, £15.95.

A QUESTION for male readers (with whose response women will probably easily empathize): how would you feel about allowing an adult male baboon to reach for your scrotum? At the very least, permitting the act would rest on a very high degree of trust; and the observation that male baboons incorporate such actions into their greeting behaviour has been interpreted as a mutual confirmation of exactly that: trust. These animals are often totally dependent on each other to maintain a coalition against others in their group but, lacking the ability to speak an oath of allegiance, they might be considered to go even further, literally putting their precious reproductive potential in the hand of the other. (It is no accident that 'testimony' shares a linguistic root with 'testicle': biblical oaths were sworn with a hand under the other person's loins.)

Such intimate trust between friends is a far cry from the popular image of these primates as ruthless, self-centred aggressors. Yet this new observation is consistent with a now substantial body of evidence that, as Frans de Waal convincingly shows, is beginning to reveal the many evolutionary foundations of human morality. De Waal disagrees profoundly with the popular view that our morality represents the triumph of cultural constraints over the self-seeking beast within us. For one thing, the pronouncements of sociobiologists such as Richard Dawkins ("if you wish, as I do, to build a society in which individuals cooperate generously and unselfishly towards a common good, you can expect little help from biological nature") and George Williams ("we need all the help we can get in trying to overcome billions of years of selection for selfishness"), are argued to be fundamentally misplaced. That genes are inherently selfish does not mean that individuals must be so. To the contrary, de Waal amasses an impressive array of evidence that animals can genuinely care about the welfare of their companions, even to the extent of showing a "community concern", which is perfectly capable of being favoured by ordinary processes of natural selection.

De Waal achieves two important things in this book. First, he successfully sets out the many ways in which the behaviour of certain species of animal reveals what he calls "the requisites of morality". A crucial ingredient here is the wealth of groundbreaking studies of monkey and ape societies conducted by the author himself over the past 20 years. He describes substantial

quantitative studies but also has an eye for the larger, interesting patterns and interpretations that may pass others by. Relying heavily on this database, but supplementing it with the findings of others on various taxa, he deals in turn with evidence for the bases of morality in animal behaviour, including empathy and sympa-



Community concern: a snow-white newborn in its mother's lap, surrounded by other stump-tailed macaques.

thy, reciprocal obligations, maintenance of group harmony and — of particular relevance — signs that animals may take account not merely of how their companions do behave but also of how they ought to behave.

De Waal's second important achievement is to say all this so well that a wide audience will find his book eminently readable. (Extensive technical notes and citations are provided in separate addenda.) Simplistic notions that we have an innate animal kernel that is inherently immoral, or that our moral sense sets us completely apart from nature, abound in writings from psychoanalysis to religion, from tabloid press to cultural anthropology. If many of those espousing such views find they cannot put the book down until they have assimilated much of the evidence, the level of debate should potentially be raised enormously. This is not to say that de Waal's book will end the debate: but it could greatly inform the discussion by laying out a profusion of observations and experimental results that are not widely known, or which have not been

brought together and interpreted as they are here.

I do, though, have one disappointment. De Waal does not offer even a sketch of a systematic evolutionary account of morality. What moral (or pre-moral) capacities are shared by the African apes, for example, and might therefore be inferred in our common ancestor? Similar questions can be addressed to wider taxonomic groups, and so on to older ancestral states. Instead, de Waal skips paragraph to paragraph from apes to jackdaws to elephants and back to primates. This makes the book entertaining reading, but gives it the flavour of a preliminary skirmish into

the fuzzy question of whether *some* animals have *some* of the (pre)requisites for morality, without suggesting where the phylogenetic differences among animals lie.

In attempting to reconstruct the evolutionary precursors to human morality, we would have to home in on the common ancestor shared between ourselves, the chimpanzee and the bonobo. The capacities of those two species would therefore need to be incorporated into the evolutionary modelling process. But de Waal concentrates more on the chimpanzee, which therefore takes on the mantle of an implicit referential model for the most recent, pre-human ancestral state. This reflects not only the emphasis of his own research but also the fact that bonobos are in general much less well understood than chimpanzees. But even the main lines of what we do know receive scant attention. The bias is surprising because de Waal has himself studied bonobos. It leaves unanswered the question of what moral antecedents we should be inferring by taking into account the repertoire of both

of our sister genera.

So a skirmish it may be, but it is a challenging and stimulating one that deserves to be widely assimilated and discussed. It may well prove to be an influential landmark in what is starting to look like a pendulum swing, from an era of social evolutionary modelling obsessed with selfishness and Machiavellian manoeuvring, to one that seeks to incorporate evidence of real niceness and community spirit, without a return to naive group-selection theory. □

Andrew Whiten, currently at Emory University, Atlanta, Georgia, USA, is in the School of Psychology, University of St Andrews, St Andrews, Fife KY16 9JU, UK.

Why we age

Roger Short

Cheating Time: Science, Sex and Ageing. By Roger Gosden. Macmillan: 1996. Pp. 339. £16.99.

THE blue-rinse set must hope against hope that age will not weary them, nor the years condemn them to Shakespeare's Seventh Age of second childishness and mere oblivion, sans teeth, sans eyes, sans taste, sans everything. Fear of ageing has haunted us forever. A popular scientific account of the biological basis for ageing should therefore be welcomed by an ever-growing proportion of the population.

There can be little doubt that lifespan is both one of nature's design constraints and subject to selection. Mike Rose's intriguing experiment with the fruitfly *Drosophila* shows how easily and rapidly increased longevity can be selected for, so we must ask ourselves what are the selection pressures that have led to such enormous differences in lifespan between species? Why have the tortoise and the hare chosen to run such very different races? Unfortunately, Roger Gosden does not really explore this problem in any depth, even though he opens with an account of *Antechinus stuartii*, Australia's spectacular marsupial shrew, where all males die at the end of each mating season, broken down by sex rather than age, because celibacy or castration will grant them an extra year of life. Ecologically it is not difficult to see method in their mating madness, but why should the rock-fish and the raven, the tortoise and the elephant, the parrot and people live such long lives? There is much that we do not understand.

The human menopause, one of the central themes of this book, also deserves more discussion. Gosden rather lamely concludes that "it is more likely to have crept on to the stage by default". He finds it difficult to see any positive advantage in

it, regarding it as "more a handicap than a benefit". Perhaps he would profit from reading Germaine Greer's alternative viewpoint in *The Change*. Because humans have the longest period of childhood dependency of any animal, it surely makes good sense that the mother should stop producing children well before she is likely to die, so that her lastborn is not seriously disadvantaged by her death. And in many traditional societies, grandmothers are not expected to bear children because they must assume a new role of caring for their grandchildren. For many women, a new life begins at 50. The shortened human female reproductive lifespan also has important consequences for our mating system. When older men remarry, they almost invariably choose a younger, pre-menopausal wife. The net result is that there are more men who father children by more than one wife than there are wives who have children by more than one husband. In other words, communities that practise serial monogamy are effectively polygynous. This may explain the persistence of such marked sexual dimorphisms in height and body build.

Perhaps the most interesting part of the book, and certainly the most readable, is the historical account of the development of 'rejuvenation therapies', starting in the late nineteenth century with the vain attempt of the 72-year-old Professor Charles-Édouard Brown-Séquard in Paris to restore his lost youth by injecting himself with aqueous extracts of guinea-pig and dog testicles. With a whimsical twist, Gosden even dedicates his book to Brown-Séquard, who at least was objective enough to realize that his therapeutic discovery might be an illusion (it was), so he offered to supply his extracts, at his own expense, to any physician who cared to investigate his claims; within a year, 1,200 had taken up his offer and were trying it on their patients. The results were unconvincing, and the method soon fell into disrepute, but it was probably Brown-Séquard's spectacular claims that drew people's attention to the potentially powerful effects of the body's glandular secretions, later called hormones, leading to the development of that eminently respectable scientific discipline, endocrinology.

The scene then shifted from Paris to Vienna, when Professor Eugene Steinach claimed that the body's own production of testosterone was stimulated by vasectomy (it wasn't). Even the poet W. B. Yeats was Steinached at the age of 69 because of his declining productivity. Vasectomy then gave way to testicular grafting, and Serge Voronoff, a French surgeon of Russian Jewish extraction, and his American counterpart, Max Thorek, started a fashion for testicular transplants, politely known as 'monkey glands'. The large testicles of chimpanzees made

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS
REASONS

Serge Voronoff: exponent of the testicle.

it possible to treat several men per animal, the grafts being placed in the scrotal sac of men seeking rejuvenation. With the recent isolation of a HIV-1-type retrovirus from wild chimpanzees, one wonders how close those thoughtless surgeons came to initiating an HIV pandemic immediately after the end of the First World War. It would be interesting to know what Voronoff's and Thorek's many patients eventually died of — perhaps their elderly and debilitated state was a blessing in disguise.

Today we tend to see testosterone more as a poison, responsible for aggression and excessive muscular development, although curiously Gosden regards such claims as unproven. Certainly, castrated men, like *Antechinus*, seem to live significantly longer, but the therapy has never caught on.

One of the most intriguing speculations about human ageing is in Aldous Huxley's *Brave New World*, which Gosden does not mention. Huxley advocated the use of "soma", a drug designed for the elderly which would be extremely pleasant to take, even addictive, but which would greatly shorten life expectancy — rather like tobacco. As countries move inexorably towards population stabilization, the age pyramid will become more columnar in shape, with ever-increasing numbers of elderly pensioners, living off the economy rather than contributing to it. Those in developed countries will be able to afford expensive palliative care, such as cancer chemotherapy or organ transplantation; most of their lifetime's expenditure on health will occur in the last few years of life, to buy themselves a little extra time at great cost. If life expectancy could be reduced by taking

"soma", there would be enormous financial savings on pensions, freeing capital for re-investment in the economy. Voluntary euthanasia is another alternative, but it is unlikely to find wide acceptance other than in the terminal phases of a chronic illness, for like Julius Caesar, we would probably all wish that our deaths might be unexpected.

If gradual senescence is our destiny, then economic incentives will make gerontology the pre-eminent medical specialty in developed countries. Meanwhile, nostrums that claim to prolong and improve the quality of life will continue to plague us. Gosden's book failed to anticipate the current US craze for melatonin as an anti-ageing hormone, based on the slenderest of evidence from animal stud-

ies. Sadly, pseudoscientific works such as *The Melatonin Miracle* by Walter Pierpaoli and William Regelson and *Melatonin: The Anti-Ageing Hormone* by Suzanne Levert have become bestsellers in US health food stores, and the Food and Drug Administration seems unwilling, or unable, to protect the gullible American public from gross misinformation and exploitation. Gosden's first foray into popular science writing about ageing will not compete with this pulp market, but at least his scientific integrity remains intact, for he has tried to tell the truth to a wider audience. In this, he has succeeded. □

Roger Short is in the Department of Obstetrics, University of Melbourne, 132 Grattan Street, Carlton, Victoria 3053, Australia.

to cool themselves nowadays?"). In evolution, faculties often appear with one function, then get co-opted for something completely different.

Other logical flaws mar Dunbar's thesis. He assumes (with no hard evidence) that group size increased in lockstep with the brain, so one can actually pinpoint the time (500,000 years ago) when growing numbers forced language to emerge. But a correlation holding within the narrow range of brains below 400 cm³ does not necessarily apply when that size is quadrupled. Dunbar sees that additional motivation for group-size increase is needed. Wisely, he drops two shaky possible explanations — protection from predators, protection from members of the same species — only to adopt a third: our ancestors became nomads. But he fails to show that nomadic primates have larger groups than others, and ignores the fact that our ancestors were not, until very recently, nomads. Migration does not equal nomadism, and caves in China were inhabited continuously for countless millennia.

But language itself holds the worst pitfalls. No-one would dare to write about, say, the evolution of sex without having spent years studying all aspects of sexual reproduction. Why, apparently, can such an apprenticeship be so freely dispensed with when the topic is the evolution of language?

Hayden Corner

Blithely ignoring the many arguments against continuity between language and animal communication systems, Dunbar assumes that language developed seamlessly from primate vocalizations, which, when "a more efficient mechanism for bonding was needed... began to acquire meaning". How was this amazing feat accomplished? Dunbar seems not even to realize how amazing it was. When he gets more specific, he commits four solecisms in one sentence. New Guinea's language is not "Pison", but Tok Pisin; BEV is not limited to "the north-eastern US"; "nation language" refers to ancestral languages of Caribbean islanders, not those they speak nowadays; and Krio, as its name suggests, is a creole not a pidgin.

The book's jacket illustration by Hayden Corner nicely captures the current landscape of language evolution studies: isolated mesas of fact separated by yawning canyons of speculation. To cross the latter requires ropes of tightly woven argument, sadly lacking here. It helps to know a bit about language, too. □

Derek Bickerton is in the Department of Linguistics, University of Hawaii at Manoa, Moore Hall 569, 1890 East-West Road, Honolulu, Hawaii 96822, USA.

I chat, thereby I groom

Derek Bickerton

Grooming, Gossip and the Evolution of Language. By Robin Dunbar. *Faber*. 1996. Pp. 209. £15.99, \$22.95.

EVER since we discovered we were apes, there has been pressure to explain human behaviour in terms of our primate ancestry. Surely, more of what we do is apelike than many would like to think — but is all of it? Language, arguably our best trick, offers slippery terrain for anyone committed to gradualist, unpunctuated views on evolution.

Undeterred, and armed with impressive credentials, including professorships in both psychology and biological anthropology, and field research on primates, Robin Dunbar confidently attacks the knotty issue of language origins. In breezy, insouciant prose, aiming at a nonspecialist audience, he proposes that language began not as a how-to manual for hunters but as a way of cementing alliances when our ancestors' social groups reached an unwieldy size.

Recent studies show how alliances between individuals play crucial roles in primate societies, counteracting the tyranny of alpha animals. These alliances are based on mutual grooming, but grooming takes time, up to 20 per cent in some species. As hominid groups grew, Dunbar claims, more alliances were needed, so more time was spent on grooming. At some point between 20 and 40 per cent, grooming would have interfered with other vital functions. Language would then have evolved as a grooming substitute, enabling our ancestors to establish and reinforce bonds with several

individuals simultaneously.

Dunbar supports this ingenious and original proposal with cross-species correlations between brain size and group size. Primates living in larger groups have bigger brains, perhaps because they need them to keep track of more complex social networks. Statistics from several

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

sources point to an optimal human group size of around 150 individuals (three times that of any other primate), making a grooming substitute indispensable. Further support is drawn from studies showing that most conversation consists of gossip about one's own or one's acquaintances' activities — a kind of verbal grooming.

But here Dunbar commits the Fallacy of Most Frequent Use. By this reasoning, computers were invented to play video games or surf the Net, while insects grew wings in order to fly ("A cooling device? Nonsense! How often do insects use wings

Chewing it over

John R. Krebs

The Natural Science of the Human Species: An Introduction to Comparative Behavioral Research — the "Russian Manuscript" (1944–1948). By Konrad Lorenz. Edited by Agnes von Cranach. Translated by Robert D. Martin. MIT Press: 1996. Pp. 337. \$35, £29.50.

THE Nobel prizewinning ethologist Konrad Lorenz wrote a 750-page manuscript while a prisoner of war in Armenia, between 1944 and 1948. It was an heroic effort. For at least part of the time, he was living in an unheated, poorly lit building (the icicles used to grow horizontally because of the wind blowing through) and his writing materials consisted largely of cut-up pieces of cement sack and potassium permanganate. He wrote with a steel-nibbed pen or home-made feather quills and his only reference work was a copy of Goethe's *Faust*. On release from prison, Lorenz persuaded the Russians to allow him to take a copy of the manuscript home with him to Austria. In his subsequent career, first in Vienna and later as director of Max Planck Institutes in North Germany and finally in Bavaria at Seewiesen, south of Munich, he used material from the "Russian manuscript" in his writings and lectures, but he did not publish it in full (in fact it was the first of what he intended as a two-volume treatise). In the early 1960s, the manuscript disappeared, and resurfaced only in 1990, after Lorenz's death, when his daughter discovered it in the library of the family home at Altenburg near Vienna. With the dedicated help of Robert D. Martin, the manuscript has now been translated into English and published.

Can a book published more than half a century after it was written be more than an historical footnote? In many ways, this volume is principally of interest for its insight into Lorenz's thinking at the time. All the main ideas in it were published subsequently in one form or another, and many of the details and the style of analysis now seem very dated.

Nevertheless, Lorenz's treatment of what he viewed as the central problem in his research career has stood the test of time. For most students of animal behaviour, Lorenz is known for three things: the development of the concept of 'imprinting', in which early experience influences later social and sexual preferences; the comprehensive "theory of instinctive behaviour"; and the use of behaviour as a tool in phylogenetic analysis. It may therefore come as a surprise that the Russian manuscript is about none of these things, although it refers to all of them. Lorenz's core theme is an evolutionary interpreta-

tion of the philosophical writings of Immanuel Kant. Perhaps because he occupied Kant's chair of psychology at Königsberg just before the Second World War, Lorenz was interested in explaining the Kantian concept of the *a priori*, the apparent built-in predispositions of the human mind that allow the individual to interpret the world outside.

Lorenz's basic question is simply this. There are real objects and events in the outside world and there are human experiences of these objects and events: how are these two connected ("the relationship between the real and the phenomenal world")? According to Kant, there were two distinct worlds: the "real world

referred to in the title (English translation: "Behind the mirror") is the human perceptual system and the "backside" is the set of evolved physiological mechanisms that enable the perceptual system accurately to anticipate and represent the outside world in a schematic way.

This evolutionary interpretation of the human perceptual system resonates well with much current thinking. Indeed, Lorenz showed remarkable foresight in anticipating the neo-Chomskian view of innate schemata for language acquisition and many other parallel ideas in what is today called 'evolutionary psychology'.

One may wonder why it took 750 (admittedly small — A5) manuscript

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Lorenz: showed remarkable foresight in anticipating many ideas in evolutionary psychology.

outside" and the internal world guided by *a priori* structures that influence perception. He did not find any logical connection between the two. ("For Kant, there was no comprehensible, logically tangible relationship between the objective world and the intelligible world comprising only those *a priori* schemata that create images of objects in our perceptual world.") Lorenz, however, recognized that there is a connection. Where do the *a priori* schemata come from? They are the product of evolution: the brain "evolved with a structure permitting it... to construct a model-like, simplified representation of external reality... determined by... inputs from the outside world". In other words, the *a priori* are evolved representations of reality, adaptive predispositions that allow the developing human to make sense of the unending, unpredictable, diverse onslaught of stimuli from the outside world. Lorenz had, of course, developed essentially the same idea for nonhuman animals when he wrote of "innate releasing mechanisms" that allow animals to respond appropriately to key stimuli the first time they are encountered in life (the baby gull pecking at its parent's bill to elicit food regurgitation, for example).

The book *Die Rückseite des Spiegels*, published by Lorenz in 1973, covers essentially the same ground. The "mirror"

pages to develop this idea. The answer is that Lorenz covers many other topics, including an overview of his methods and results of studying animal behaviour. He is characteristically vivid in his metaphor and illustration. He demonstrates the lack of pain in insects by recalling "the grotesque sight of a grasshopper voraciously devouring its own rear end". On consciousness: "While this book is being read, the reader's pyloric reflex is alternately opening and closing the outlet from the stomach and thus delivering to the duodenum finely gauged quantities of predigested food pulp. The reader has just as little experience of this process as does an amoeba of its highly appropriate responses to a prick from a needle."

I suspect that few will wish to read the book from cover to cover, with or without their pyloric sphincter in action. But for those interested in Lorenz and his work, there are some quotable quotations and thought-provoking insights into a man whose creativity and influence on his subject area in the middle part of this century were immense. □

John R. Krebs is at the Natural Environment Research Council, Polaris House, North Star Avenue, Swindon SN2 1EU, UK and Department of Zoology, University of Oxford, Oxford OX1 3PS, UK.

And so to the bedside

Roy Porter

Becoming a Physician: Medical Education in Britain, France, Germany and the United States 1750–1945. By Thomas Neville Bonner. Oxford University Press: 1995. Pp. 412. £30. \$35.

THOMAS Bonner has a complex story to tell. This is partly because, as the German physician Johann Heinroth observed at the beginning of the nineteenth century, “the Italian loves the old, the Frenchman the new, the English solid ground under their feet, and the German all of this”. And alongside different national traditions, medical education had to cater for all sorts, from future royal physicians to country doctors, midwives and bath-keepers. Small wonder, therefore, that it assumed many different forms, from humble apprenticeship to illustrious universities.

And the diversity in emergent medical schools was increased by two further factors. On the one hand, social rank and pretension — most tellingly reflected in whether instruction was to be in Latin or the vernacular. And on the other, variations in national political climate — *laissez-faire* England adopted private schools, Prussia pursued authoritarian centralism, pre-Revolutionary France empowered faculties and corporations. One great strength of Bonner’s superb book is its grasp of the many and distinct strands that have gone into the skein of medical education.

Another virtue of this rarity, a genuinely comparative study, is Bonner’s eye for change. Innovation was in the air during the late Enlightenment. In many teaching centres, chemistry, physiology and botany became more prominent, while burgeoning anatomy schools taught practical dissection and guided the knife of pathology. Hospitals assumed new importance as sites of medical instruction with the invention of clinical lectures. Not least, the training of physicians and surgeons drew together.

Such piecemeal and informal innovation was given a new cohesion by striking reforms brought in by the French Revolution. Radical reorganization turned public hospitals (for the poor, of course) into giant arenas of clinical teaching. For the next half century it was easier for a student to gain access to a patient’s bedside (or later to cut up his corpse) in Paris than anywhere else in the world; and that is why students from the Old and New Worlds alike flocked there, eager to learn what was called “practical medicine”,

above all gaining diagnostic expertise based on pathological anatomy. But the resulting crush of student numbers proved counterproductive, which partly explains the growing attraction in later decades of the smaller German polyclinic, attached to a university, where everything was more intimate, hands-on and supervised.

Assessing such changes, Bonner suggests that student ‘demand’ played a large

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Sir James Young Simpson invigilating the MD examination at Edinburgh in 1855.

part in provoking them. And certainly numbers were sufficient to exert real consumer pressure: there were well over a thousand medical students in Paris around 1800 and all could shop around. Nor were medical students slow to voice their grumbles, here against lack of cadavers to dissect, there about the hodgepodge of courses that passed for an education. Not least, the nineteenth-century medical student was extraordinarily mobile, free and eager to travel to wherever the best (or most renowned) instruction was on offer. In the 1820s, morbid anatomy in Paris was the great draw; by 1880 that had been supplanted by histology in Heidelberg or bacteriology in Berlin. And why had Germany taken over? Bonner suggests that the tremendous success of medical training on German campuses after 1850 was due to inter-regional competitiveness combined with government policies of pedagogic paternalism that viewed generous funding of higher education as a national investment.

No less powerful a factor than ‘demand’ in promoting nineteenth-century reforms, however, was the desire of the authorities

in London, Berlin or Dublin for ‘order’: rowdy and ribald medical students were notoriously disruptive, catcalling their lecturers and engaging in pranks and orgies of public drunkenness. The systematization of state-regulated licensing and the invention of the formal medical school, with its deans and dinners, rowing and rugby, were in large measure an attempt to curb and channel the disruptive student culture, feared in Germany in particular as a hotbed of radicalism. Tighter regulation of medical licensing was also a response of governments, in increasingly democratic times, to the need to allay public anxieties through bolstering professional legitimacy.

Bonner also offers a convincing explanation for what was probably the most important transformation in twentieth-century medical education, that following the 1910 Flexner report, which condemned so many medical schools in the United States as lamentably wanting. There is no doubt that students and professors alike welcomed the new accent on basic science that the Flexner reforms brought. But the report’s author and champions were no less motivated by the wish to outlaw the proliferation of colleges catering for fringe sects, for women and for blacks. The Flexner reforms were a triumph for science, but they were no less a prop for the regular profession.

Much else is considered in this book besides, including the struggle of women for entry into the profession. And, looking bottom-up as well as top-down, Bonner has an ear well-tuned to the perennial whinges of the medical student: no money, cold digs, bad food, indigestible lectures and “half of the time we never get near a bedside”.

Considering its importance, the history of medical training has been oddly neglected, or studied mainly as institutional biography. Bonner’s account is far richer, embracing supply and demand (lots here on what medical students actually wanted), curricula and experience. Taking the long and broad view, he achieves proper historical distance from those pedagogic nostrums that, over two centuries, have successively been advocated as panaceas for the maladies of medical education: a structured curriculum, formal examinations, bedside work, basic science, more bodies, more laboratory work, more casework, and so on up to the present. His verdict is that there seems no royal road to ‘becoming a physician’. Clear, concise and comprehensive, this study will long remain definitive. □

Roy Porter is at the Wellcome Institute for the History of Medicine, 183 Euston Road, London NW1 2BE, UK.



THE brittle star *Ophiolepis impressa*, a graceful agile relative of the sluggish sea star. Taken from *Sea Stars, Sea Urchins, and Allies* by G. Hendler, J. E. Miller, D. L. Pawson and P. M. Kier, a comprehensive guide to the identification and natural history of the echinoderms of Florida and the Caribbean. Smithsonian Institution Press, \$39.95.

Invisible strangers

June Goodfield

Women Scientists in America: Before Affirmative Action 1940–1972. By Margaret W. Rossiter. Johns Hopkins University Press: 1995. Pp. 584. \$35.95, £30.

TWELVE years ago, the first volume of Margaret Rossiter's marathon opus, *Women Scientists in America*, was published to great praise. It was quickly recognized as a landmark study, and most reviewers, myself included, found their appetites whetted for the next volume, which would take the record from 1940 to — we thought — the present day.

Well, the next volume has now appeared. It takes the story a further 32 years to 1972, and is some four-fifths as long as the previous one. And we are probably going to have to wait at least another 12 years before the story is brought up to the present day, by which time the century will have turned.

As I said in my previous review (*Nature* 302, 761; 1983), Rossiter gave us "sophisticated analysis of a complicated situation encapsulated in this phrase: 'it is the history of an occupational group whose status had risen and fallen over time as the women's role responded to external events and pressures'". I ended by asking: "Beyond the present who can say what new pressures may influence the status of women in science, whether in Europe or in America?" Reading this next volume, one is tempted to answer: "the same old ones!"

The first volume provided a story whose *dramatis personae* contained coura-

geous pioneers as well as obscurantist reactionaries — and men and women were both found in each category. The movement to educate women — whose real objective was the production of more enlightened sons — created "a group of highly motivated, qualified and quite remarkable people". But the result, unfortunately was not a recognized cadre of professional women accepted on a par with their male scientific colleagues but a cadre of highly qualified women who had no place to go except back to where they came from — the women's colleges.

Two important things then happened in relatively quick succession, historically speaking: the Second World War, followed by 32 years recognized as a golden age for science in the United States. And under the influence of Vannevar Bush and the National Science Foundation and the National Institutes of Health, science went on to flourish as it had never flourished before, measured by whatever criteria one might choose, whether money spent, persons trained, jobs created, articles published or Nobel prizes won. Yet, as Rossiter asks in her introduction, if this was so, why were women still invisible?

At the beginning it had seemed all so promising. The Second World War was a complete and tragic disaster for many people in many countries, but it did galvanize the status of women and was probably the then single greatest agent of social change. In the United Kingdom, for instance, women joined the Women's Royal Naval Service or the Auxiliary Territorial Service or the Women's Royal Air Force; they manned anti-aircraft guns; they flew aeroplanes across the Atlantic; they made munitions in factories; they dug

the land and harvested the crops; and they joined — as scientists — Britain's radar research establishment in Malvern. And when the war was over, most did not want to return to the confines of home and family, or to a life of domestic help, serving the aristocracy or the *nouveau riche* of the middle class. They even became strangers to their husbands, as the delectable film *Perfect Strangers*, featuring Deborah Kerr and Robert Donat, beautifully showed.

The war changed social attitudes and behaviour in the United States too. Horsemeat, initially offered on the Friday menu of Harvard University's Faculty Club so that there was one 'non-meat day' a week in the interests of austerity, was rapidly removed when the war ended but quickly reappeared after cries from an outraged faculty. (How unlike the dear common-room life at Oxford and Cambridge!) But, as Rossiter points out in her admirably clear introduction, the war apparently saw the start of a remarkable period for women in science. They were told that they could do anything; they were recruited for certain scientific and technical projects; and the manpower demands of a highly technological military-industrial complex saw officials launch a campaign for 'women power' as they urged bright women to train in the nontraditional areas of science and engineering. As a result, record numbers of women earned doctorates in scientific and technical fields, but they disappeared. Why?

Rossiter sets out a series of questions she then proceeds to answer. Did the specific areas or fields of study in which these women operated happen to grow less slowly than those of men? Did the scientific job market work differently for them? Did marriage or marital status in general have a limiting impact on their careers? And, if the answer to all these questions was yes, why and how did these limits to women's opportunity rise and fall?

Once again, Rossiter displays an intriguing and yet infuriating paradox. In the first volume she showed how all the efforts to educate women took them back into their own colleges. In the second volume she shows that, against all expectations, the period under questions was a Dark Age for women in all the professions, not only science:

The growth and affluence of the period that could have made room for more and better trained scientists of both sexes, did not benefit the two equally; in fact, they generally unleashed certain forces that hastened the women's exit and subsequent marginalisation and under-utilisation, which could then be cited to justify denying further training for their successors.

Perhaps most tragically of all — and this could even be considered a form of

betrayal — women's colleges, far from being the benefactors to their kind, threw up the greatest obstacles. Rossiter presents convincing evidence that:

Most of the women's traditional employers, such as women's colleges, teacher colleges, and colleges of home economics, were closing their doors... and they [women] were not included on the faculties of the growing and new co-educational institutions. Much of this new exclusion was tied to marriage, as in the reinstatement of the ante-nepotism rules on most campuses after World War II. But even single women were ousted or just not hired in the pronatalist years, often for fear that they might later get married.... Thus the pattern was deliberate and grew widespread.

Matters were to change of course, but it took at least 20 years before even a wedge was insinuated into the closed door — although 'drawbridge' and 'portcullis' might be better metaphors. For not only did men return from the war and re-establish themselves in the hierarchy of the professions, with the naturally accepted attitudes of dominance, but recruitment in universities and liberal arts colleges, previously female-dominated areas, now became masculinized. Perhaps most chilling is the way in which this change was more than tacitly endorsed by foundation officials and academic administrators. So in a time that gave a young male scientist every opportunity and enhanced status, young women were "supposed to be at home with the children, whether they had them or not, or whether they wanted to be there or not".

Discrimination was a word that did not come into use until the late 1960s. Before then the social patterns were seen as just the way things were, always had been and should always remain. So those professional women who clearly understood what was happening, who did see a pattern and reported it, were understandably reluctant to criticize the powerful and successful. Even those who actually tried to do something, to correct the situation, were more often than not ineffectual. Some were ineffectual because of the strength of the opposition: Margaret Mitchell's article in 1951 on these matters was shot out of the sky by a counterblast from a Harvard psychologist of such force and venom that the topic was never raised again for a decade. Sometimes they were ineffectual through their own muddled thinking and passion for compromise, as when a committee of Harvard Faculty and Radcliffe Trustees wrote a report on the limits of women's opportunities in the academic world that was so bland and compromising it was totally counterproductive. Yet such was the prevailing mindset, and social ambience, that, as Rossiter claims, if the data *had* been interpreted to demonstrate discrimination, they would

probably not have been published; if they had been published, more likely than not they would have been ignored.

When the change came, it came not from within science but from the activities of many social scientists set against a larger social reality. And there was now one new element within this larger reality that had nothing to do with the science.

The civil rights movement proved to be the agent. Rossiter argues that Alice Rossi, who devoted many years in the early 1960s to rethinking and reformulating prevailing wisdom, was able to come up with a new view only because, in the context of the civil rights movement, she could really *see* a pattern. True, others had too, but she could delineate the complex of oppressive attitudes and practices. Women in science did not deserve their fate and should not be blamed for their obscurity, in the same way that blacks did not deserve *their* fate and could not be blamed for it, either. Society could and should be changed; laws would have to be passed by Congress, however reluctant, and finally the executive branch would have to be pressured into enforcing the laws. By 1969, the anger of women had coalesced into a movement; innumerable reports on the status of women followed, angering even more women as the totality of their exclusion was documented further and brought fully into view. By 1970, federal hearings on sex discrimination on the campus and workforce were being held; by 1972, landmark legislation was in place on equal pay and affirmative action in

academic institutions.

This then is the general thesis for which the 16 chapters in this book provide irrefutable evidence, assembled with the careful scholarship that has become Rossiter's hallmark. Once again, the quantity of material researched is enormous, and my caveats are few. So much is involved in these 20 years, so much to be covered, that I found the book heavier going than its predecessor, with fewer of the light touches of irony and humour that I had enjoyed so much before. Rossiter declares in her final sentence that these 32 years marked "the ending of an era and the beginning of a new and more equitable one". Yet those who rejoice and applaud what has happened should always reflect that, as George Steiner once wrote about science, there will always be moral ambushes waiting for us. Nothing will be clear-cut and simple. Reading social history, whether George Trevelyan's, Asa Briggs' or Margaret Rossiter's, I am pulled back time and again to what William Blake wrote in *The Vision of John Bull*: "I ponder on how men fight and lose the battle and the thing they fought for comes about in spite of their defeat. And when it comes turns out to be not what they meant and other people have to fight for what they meant under another name." □

June Goodfield, emeritus professor at George Mason University, Fairfax, Virginia, is at International Health and Biomedicine, The Manor House, Alfriston, East Sussex BN26 5SY, UK.

Sceptical wonderment

Alan Cromer

The Demon-Haunted World: Science as a Candle in the Dark. By Carl Sagan. Random/Headline: 1996. Pp. 436. \$25.95, £18.99.

CARL Sagan's high "wonder quotient" was tempered at an early age by the scepticism of Martin Gardner's *Fads and Fallacies in the Name of Science*. And wonder and scepticism have been "uneasily cohabiting" modes of thought throughout Sagan's career. Both he and Gardner have long been affiliated with the Committee for the Scientific Investigation of Claims of the Paranormal, publisher of the *Skeptical Inquirer* (SI), and Sagan writes as a committed, indeed, a crusading, sceptic about many of the incredible beliefs familiar to SI readers. There is much fun to be had in reading about the debunking of such 'mysteries' as the English crop circles (produced by a pair of dedicated pranksters) and the crash in Roswell, New Mexico, of an alien spacecraft (a secret high-altitude

military balloon).

Sagan writes regularly for *Parade*, a Sunday newspaper magazine that reaches a staggering 37 million US households (83 million readers). Seven chapters in this book are an expansion of a 1993 *Parade* article on why he does not believe in alien abductions. Other chapters were written on different occasions for different audiences, making for much repetition and confusion of purpose. Some are clearly for students in his Cornell University course on critical thinking, some are based on articles in *Parade*, some are for scientists, and the final chapter is based on an address given at an induction ceremony for new US citizens.

Sagan is a forceful advocate for science and a fierce opponent of pseudoscience, mysticism and religion. "Science", he writes, "is different from any other human enterprise... in its passion for framing testable hypotheses, in its search for definite experiments that confirm or deny ideas, and in the vigour of its sub-

stantive debate, and in its willingness to abandon ideas that have been found wanting." But as good as this sounds, it fails to acknowledge the central importance to science of a theoretical framework. Of all the many claims of the paranormal that Sagan catalogues, only perpetual motion is rejected on theoretical grounds; all the others are rejected for lack of evidence. Thus, both in theory and practice, Sagan takes an essentially empirical-inductivist view of science.

This leaves him combating pseudoscience with one arm tied behind his back, for without a positive theoretical reason ("We know things don't work that way!") the mere lack of evidence is a weak argument against belief. No matter how many alien abduction scenarios and UFO sightings are explained away, one can never answer the reasonable question of one of his *Parade* readers: "A hallucination might account for 99% [of reported alien abductions], but can it ever account for 100%?"

Since at least the time of Archimedes, science has deduced results from very general principles, which are not themselves capable of direct verification. The Greeks, for example, deduced the law of reflection from the principle that light should travel the shortest distance from the eye to the object. From the law of reflection, they deduced the images formed by plane and curved mirrors.

This is difficult stuff, and requires much training and study to learn. Sagan says the same about modern science at several points in the book, but in one chapter he vehemently contests the view that science is not part and parcel of human nature. Science is a natural proclivity, he asserts, that "is embedded deeply within us". Science is not hard to teach because it is hard, but because "indifference, inattention, incompetence, or fear of scepticism... discourage children from science". Why there is so much incompetency and fear of a natural proclivity is not explained, although elsewhere he admits that seven-year-olds are much better at wonder than scepticism.

The innateness of science is a deeply held Sagan belief. He denies that science had a unique origin in ancient Greece, claiming that the tracking know-how of preliterate peoples is science comparable to that of the Greeks. This claim goes beyond rejecting the theoretical accomplishments of the Greeks, to rejecting his own views of science given elsewhere in the book. If tracking know-how is science, why not make the same claim for a chimpanzee's ability to collect termites with a grass reed? Amazingly, Sagan does, defending his belief as uncritically as any alien abductee.

Sagan's passionate advocacy for science and scepticism uneasily cohabits with his passionate advocacy of shaky scientific, political, social and moral posi-

tions. The result is a book with an abundance of material on which an aspiring sceptic can practise. □

Alan Cromer is in the Department of Physics, Northeastern University, Boston, Massachusetts 02115, USA.

Regal recipes

John Mann

Classics in Total Synthesis: Targets, Strategies, Methods. By K. C. Nicolaou and E. J. Sorenson. VCH: 1996. Pp. 798. DM128, \$80, £52 (hbk); DM78, \$49.95, £32 (pbk).

IT is not often that a chemical structure graces the front cover of *Nature*, but such was the honour accorded to the complex natural product taxol* on 17 February 1994. It advertised the publication (in that issue — *Nature* 367, 630; 1994) of a paper recording an elegant total synthesis of taxol by K. C. Nicolaou and colleagues. The molecule had provided a formidable challenge (and still does) with its 11 stereocentres and dense array of functionality, yet the synthetic pathway reported was intriguingly simple, stereoselective and at least partially convergent. In short, a classic total synthesis and one of the 36 such syntheses gathered together by Nicolaou and Sorenson in this book.

But what constitutes a classic total synthesis? The authors have restricted themselves to natural products rather than theoretically interesting molecules such as cubane or dodecahedrane, or man-made structures such as dendrimers or spherands. So most of the examples demonstrate the great artistry and ingenuity of synthetic chemists as they struggled to assemble ever more complex natural products during the past 40 years. Much of this artistry depended on the invention of new reactions or methods for the control of stereochemistry, and perhaps most importantly on the ability to analyse the problem before starting the syntheses. E. J. Corey's exposition of the rules of retrosynthetic analysis in the late 1960s revolutionized the way chemists plan a synthesis, and his book *The Logic of Chemical Synthesis* (Wiley, 1989) summarized the methodology and reviewed his own 'classic' syntheses.

Each of the 36 chapters in Nicolaou and Sorenson's book spotlights one natural product and reveals the intimate details of its synthesis. Woodward's route to strychnine is not unexpectedly in 'poll position', not least because of its ingenuity and use of none but the simplest of chemical reagents. Although completed in 1954,

it remained the only total synthesis of this complex molecule until the newer routes of Magnus, Overman and others appeared in the early 1990s. It was also important for demonstrating that a complex molecule could succumb to total synthesis, and so encouraged others to take the plunge. Woodward was also involved, with Eschenmoser, in the 12-year international effort that culminated in the synthesis of vitamin B12 in 1973. Like so many of the syntheses detailed in the book, the invention of new enabling chemistry along the way was just as important as the final triumph, and these novel reactions are highlighted throughout the text.

Not all of the molecular targets are complex, and the synthesis of the simple monoterpene menthol by Takasago and colleagues is used to exemplify important advances in catalytic asymmetric reactions. The preparation of the simple hexoses using Masamune-Sharpless asymmetric epoxidation and related techniques provides a further insight into the power of these new techniques. But it is the big molecules (the 'synthetic Everests') that take pride of place. Synthesis of the anti-neoplastic macrolide cytotaricin provides an opportunity to explore the use of the Evans chiral auxiliaries in asymmetric synthesis; and the preparation of the immunosuppressant rapamycin involves an examination of the palladium-catalysed reactions introduced by Heck, Stille and Suzuki. The current vogue for tandem radical processes is brought to the fore in the chapter on hirsutene synthesis, and 18 other syntheses that use this strategy are also discussed. The last two chapters reveal the enormous complexity of marine natural products and chronicle the triumphs of Kishi's and Nicolaou's groups in their syntheses of palytoxin and brevetoxin B.

Along with all this methodology, each chapter includes essential information about the occurrence and biological significance of the natural products, a full retrosynthetic analysis of the target structure and a detailed account of the actual synthesis. Mechanisms of the reactions and problems encountered are also examined. The quality of the illustrations is superb and there is a comprehensive list of references in each chapter. The book is not only a record of synthetic achievement during the past 40 years, but also a valuable research and teaching aid.

Not everyone will agree with the authors' choice of 'classics'. The few European contributions, for instance, are largely buried within the various chapters. But there is no denying that this superb book about the art and science of natural product synthesis will be an essential purchase for many organic chemists. □

John Mann is in the Department of Chemistry, University of Reading, Reading RG6 6AD, UK.

* Bristol-Myers Squibb has registered Taxol as a trade mark and wishes the scientific community to use the name paclitaxel.

Long-period volcano seismicity: its source and use in eruption forecasting

Bernard A. Chouet

At an active volcano, long-period seismicity (with typical periods in the range 0.2–2 s) reflects pressure fluctuations resulting from unsteady mass transport in the sub-surface plumbing system, and hence provides a glimpse of the internal dynamics of the volcanic edifice. When this activity occurs at shallow depths, it may signal the pressure-induced disruption of the steam-dominated region of the volcano, and can accordingly be a useful indicator of impending eruption.

THE science of volcano seismology aims to understand the dynamics of active magmatic systems, determine the physical properties of such systems, and map the extent and evolution of source regions of magmatic energy, all of which represent critical steps in our understanding of eruptive behaviour and in the assessment of volcanic hazards. A volcano is a characteristic manifestation of the interaction of magmatic activity with the surface of the Earth. The important factors defining a volcanic structure are the duration and rate of magma transport, both of which are episodic in character as a result of the inherent instability of magmatic systems at all timescales¹. This episodicity is reflected in seismic activity, which originates in a volume of the Earth's crust hosting magma transport paths ranging from well defined loci typified by the feeding conduits and rift zones of single symmetric volcanoes, to bifurcating sets of dykes and sills underlying basaltic cinder fields, basaltic flow fields, and large silicic systems. This seismicity manifests itself in two distinct types of process, namely those originating in the fluid and those originating in the solid.

Processes originating in the fluid result in pressure fluctuations caused by unsteady mass transport and/or thermodynamics of the fluid. By their very nature these processes—and, in particular the frequencies of the pressure fluctuations—provide direct informa-

tion regarding the state of the fluid. Seismic events originating in fluid processes typically include long-period (LP) events and tremor. Long-period events resemble small tectonic earthquakes in duration but differ in their characteristic frequency range and harmonic signature (Fig. 1). Tremor is characterized by a harmonic signal of sustained amplitude lasting from minutes to days, and sometimes for months or longer. In many instances LP events and tremor are found to have essentially the same temporal and spectral components^{2,3}, suggesting that a common source process, differing only in duration, underlies the two types of event.

Processes occurring in the solid rock manifest themselves mainly in the form of earthquakes associated with shear failures in the volcanic edifice whereas LP events involve volumetric modes of deformation. The former earthquakes act as gauges that map stress concentrations distributed over a large volume surrounding magma conduits and reservoirs. They are named volcano-tectonic (VT) earthquakes to differentiate them from pure tectonic earthquakes occurring at plate boundaries and elsewhere, although they are indistinguishable from the latter in their broadband spectral characteristics.

Volcano-tectonic seismicity is often the first sign of renewed volcanic activity^{4,10}, and some eruptions have been immediately preceded by strong short-term increases in VT activity^{6,10}. Precursory VT activity may last from days^{6,10} to months^{4,5,7} or even years^{6,8,9}. Hence VT seismicity can be both a good long-term sign of potential unrest as well as a good short-term indicator.

In contrast to VT activity, which is generally more spread out in space and time, LP/tremor activity originates in particular locations within the magma plexus where disturbances in the flow are encountered. This property takes special significance at the locus of interaction between the magma conduit and the groundwater system in stratovolcanic edifices, where the internal disruption of the hydrothermal system and subsequent rapidly developing onset of volcanic eruptive behaviour over periods as short as days to weeks, provide a rich source of pressure perturbations and attendant precursory LP activity. Volcanoes in this state of pressurization are vibrantly 'alive' and pulsating.

Although we are not yet at the point in our understanding of either type of seismicity for either to be an infallible prediction tool, in this author's view LP activity shows particular promise because of its potential for understanding the physical processes that lead to its generation and their practical application to eruption forecasting. Accordingly, I will focus here on LP activity.

The work reviewed below has shown that the basic element of LP activity is the underlying pressurization of magmatic and/or hydrothermal fluids and attendant pressure perturbations, either self-excited or forced, which by their coupling to the solid radiate the characteristic signals of LP events. Crucial to the use of LP events for extracting information about the state of the fluid is our

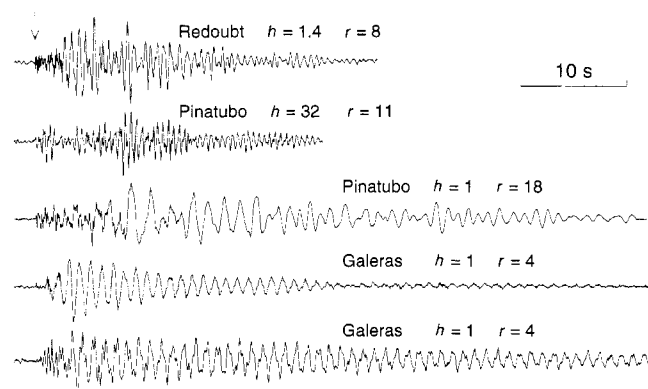


FIG. 1 Typical signatures of long-period events observed at Redoubt, Pinatubo and Galeras volcanoes. Source depths, h , and epicentral distances, r , in km, are indicated above the seismograms. The traces have been aligned according to the onset of the compressional wave marked by the arrow. The signatures are all characterized by a harmonic coda following a signal onset enriched in higher frequencies. These features are best illustrated in the shallow event at Pinatubo (middle trace), where a high-frequency onset lasting ~ 10 s is seen to merge into a quasi-monochromatic coda lasting ~ 50 s.

ability to discriminate between LP and VT events. Our studies demonstrate and illustrate that spectral analyses of LP events are needed to recognise the differences between the source properties of fluid-driven activity and those of solid-state failure processes.

Seismic signatures of volcanic activity

The identification of LP events with eruptive activity is scarcely new^{11–13}, but the significance of this activity for eruption potential was not clearly recognized until a formal theoretical background was established that linked the LP signatures to physical processes at the source. Two outstanding examples of precursory swarms of long-period events are those that preceded the 1958 and 1983 eruptions of Asama volcano, Japan^{14–16}. Swarms of long-period events also preceded the 1987 eruption of Meakan-dake, Japan¹⁷, the 1989 Ito-oki submarine volcanic eruption near the Izu peninsula, Japan¹⁸, and the 1982 eruptions of El Chichón, Mexico¹⁹. Intense swarms of LP events, some of which were large enough to be recorded at distances of up to 300 km, preceded the paroxysmal eruption of Mount Pinatubo, Philippines, in 1991⁵. Precursory LP swarms were recognized for 14 of 22 significant eruptions at Redoubt, Alaska, in 1989–90²⁰. Many additional reports of precursory LP activity can be found in the literature^{3,21–31}.

Long-period events all share a characteristic signature consisting of a high-frequency onset followed by a harmonic waveform containing one or up to several dominant periods in the typical range of 0.2–2 s (Fig. 1). The high-frequency beginning of the signal is most clear when observations are made close to the source where the effects of anelastic attenuation are minimized (see Fig. 1). An important limitation in recognizing LP seismicity as diagnostic of impending eruptive activity has arisen from the wide variety of names used to describe it. For example, in Japan LP events have been identified variously as *b*-type events^{14,32}, *N*-type earthquakes¹⁵, tremor-like volcanic earthquakes¹⁶, single-frequency earthquakes¹⁷ and isolated tremor¹⁸; at El Chichón volcano they were named type-2 and type-3 earthquakes¹⁹, and at Galeras volcano they were referred to as screw-type events³³ and butterfly-type events³⁴. Although there is no one-to-one correspondence between these events and those in Fig. 1, there is no mistaking the resemblance between these waveforms. Another impediment has been the idea that the LP signature reflects wave propagation effects along the transmission path³⁵ rather than characteristics of the source mechanism as accumulating evidence now strongly suggests.

The essential features of the LP signature become clearer when they are compared to those of VT earthquakes or tremor. For example, Fig. 2 shows seismograms and their associated spectrograms for several sources observed before the 14 December, 1989 eruption of Redoubt volcano, Alaska. The LP event is characterized by a weak high-frequency onset with frequencies up to 13 Hz, followed a few seconds later by a strong harmonic signal with a peak frequency near 1.5 Hz lasting roughly 30 s. In contrast, the deeper VT earthquake is dominated by broadband P and S phases with energy peaking in the band of 6–8 Hz, followed by a very short broadband coda (the trailing part of the seismogram) with significant energy up to 15 Hz. This signature is typical for VT sources deeper than a few kilometres and is easily distinguished from that of LP events. The shallow VT earthquake resembles the deep VT event but displays a slightly longer, narrower-band, coda exhibiting frequencies characteristic of a dispersed surface wave-train. Shallow VT earthquakes usually generate stronger surface waves than the deeper VT earthquakes but are still distinguishable from the LP signature. The hybrid event shows mixed characteristics reminiscent of both LP and VT events. The high-frequency onset of the hybrid is more pronounced than that of the LP, but its coda is dominated by a non-dispersive harmonic wavetrain that is characteristic of LP events so that the spectrograms of the codas of hybrid and LP events are similar. In terms of first motions, the VT and hybrid earthquakes both show mixed polarities at the recording stations, whereas LP events typically display the same polarity at all stations³⁶. Hybrid events can usually be distinguished from

LP events on the basis of this polarity distribution. The LP, hybrid and shallow VT events in Fig. 2 all occurred in roughly the same source region at depths of 1.4–1.7 km beneath the crater of Redoubt, yet their temporal and spectral signatures are quite distinct. As they were all recorded at the same location, their distinct character cannot be attributed to a path effect. Comparisons with the signatures of a chemical explosion detonated in the crater of Redoubt also indicate that these features can not be attributed to site effects². The tremor signal closely resembles that of the LP event except in its sustained character. The two types of event share the same dominant frequency near 1.5 Hz, and the tremor displays occasional bursts of high frequencies (see signal near 20 s) reminiscent of those found in the early part of the LP signature. These features are suggestive of a common source process, differing only in duration, underlying LP events and tremor.

Source dynamics

The above considerations naturally lead to a classification of seismic events based on the physics of the source process, which has the advantage of offering a clearer picture of the link between fluid transport and seismicity. Viewed in such a context, the enormous variety of seismic signatures observed in volcanoes is merely a reflection of the effects of extreme structural heterogeneity and strong topography of volcanic edifices on the signals originating in two basic families of processes. The first family consists of volumetric sources, in which the fluid plays an active role in the generation of elastic waves, and the second consists of shear or tensile sources involving brittle rock failure.

In volumetric sources, gas, liquid and solid are dynamically coupled and elastic radiation is the result of processes originating in the physics of multi-phase fluid flow through cracks and conduits. Long-period events, tremor and seismic signals related to mechanisms of degassing in open vents are manifestations of such processes. The liquid and gas may be of magmatic or geothermal origin depending on the volcanic setting, and the gas content may vary greatly according to source depth and gas fugacity. For example, the sources of long-period events and tremor beneath Kilauea are thought to involve magma with some gas, the proportion of which increases with the shallowness of the source^{3,37,38}. In contrast, the shallow long-period events and tremor under Nevado del Ruiz are interpreted as originating in the thermal interaction between magmatic heat and a separated groundwater system, such that liquid water and steam are the primary constituents of the two-phase fluid involved in the source process^{23,24,38}. At Galeras, shallow LP activity appears to be intimately linked to the accumulation of SO₂ beneath the crater floor³¹.

The second family of sources involves brittle failure and purely elastic processes. These include the VT earthquakes, in which magmatic processes provide the source of elastic strain energy that leads to rock failure, but where a fluid is not actively involved in the source dynamics—although fluids may be passively involved by reducing effective stress and strength through elevated pore pressures. These sources occur in the brittle rock around the magma reservoir and conduit (either within the volcanic edifice or in the crust beneath it) and may involve shear linked to stresses induced by magma movement^{29,39,40}, or tensile failure of rock caused by thermal contraction due to cooling in the vicinity of a magma body or in the volcanic edifice as a whole^{41,42}. Thus, these sources are associated with the structural response of the volcanic edifice to the intrusion and/or withdrawal of fluids. In this sense they differ from purely tectonic earthquakes, which are driven by large-scale tectonic plate motions, although their source processes and signatures are indistinguishable from those of tectonic events.

In the greater dynamical context of an entire volcanic structure, LP events, tremor and VT earthquakes are intimately related and inseparable. Thus, there must be classes of events that represent the transitions between the idealized end-member families

described above. These correspond to the so-called hybrid events seen in Fig. 2, which are thought to involve shear faulting on a plane intersecting a fluid-filled crack and thus may contain both shear and volumetric components. Consequently, there should be tremor and/or LP-induced failure cascades in which the percolation of fluids through fluid-induced fractures associated with magma transport activates a cascade of shear failures^{38,39,43,44}. If true, this means that there should be some sets of hybrid events blending the characteristics of LP events, tremor and VT earthquakes. This is supported by the fact that LP events and tremor in Hawaii volcanoes are clustered within or near clusters of VT earthquakes⁴³. This may well be true also for two-phase groundwater systems at high enough pressures to induce shear failures by hydrofracture, as suggested for the hybrid events observed at Redoubt³⁶.

An interesting aspect of LP seismicity is the similarity in the signatures of individual events in some swarms, which is strongly suggestive of the repetitive excitation of a stationary source in a non-destructive process^{2,20,27,32,36}. Another important clue comes from the observed similarities between LP events and tremor. Both have been observed to occur concurrently with the same spatial, temporal and spectral characteristics^{2,3}. Both display strongly peaked spectra with dominant frequencies that are independent of azimuth and distance to the source^{2,45}. Both show correlations between changes in their spectra and changes in volcanic activity^{3,45,46}, and both can display wide variations of spectral amplitude and corresponding small variations in frequency during steady activity^{3,30}.

The repetitive nature of LP seismicity, together with the ringing characteristics of the LP signature and the close link between LP events and tremor, all point to a source effect involving some form of resonant excitation in a fluid-filled conduit. Tremor and LP activity are thus different manifestations of the same basic process of unsteady mass transport. In this context, the LP event may be viewed as the response of the tremor-generating system to a sudden pressure transient, whereas sustained tremor may be interpreted as the response of this system to sustained pressure fluctuations. The occurrence of one or the other depends on the physical parameters of the fluid and flow conditions in effect at the time so that the richness of the signatures observed is a direct consequence of the sensitive dependence of the nonlinear fluid dynamics on these flow parameters. Recent pre-eruptive sequences observed at Redoubt^{2,20}, Pinatubo^{5,47}, Galeras^{21,25,31} and Ito-oki⁴⁸ are examples of this variability of behaviour.

Put together, these observations suggest the working hypothesis that shallow LP activity, and tremor, are manifestations of pressurization in a magmatic/hydrothermal system. Accordingly, one might expect that a direct link should exist between the strength of LP activity and the potential for explosive activity; in other words, one might expect LP events to be more likely to develop into an energetic swarm in explosive systems supporting strong pressurization, than in effusive systems for which pressurization is obviously weaker. By the same token, one might expect that the rate of LP production should depend on the rate and magnitude of pressurization. Thus, a sealed system may support strong pressurization, which may manifest itself in an energetic LP swarm, whereas a leaky system may sustain little pressurization and relatively few LP events. Observations at Redoubt in December 1989 and Galeras in January 1993 provide contrasting cases of pressurization in sealed versus leaky systems. Redoubt, which was capped by a lava dome at the time^{2,49}, is an example of a sealed system with well developed precursory LP activity. In contrast, Galeras, which had been left with a relatively unobstructed conduit after the July 1992 explosion³¹, is an example of leaky system with subtle precursory activity. At Redoubt, the pre-eruption rate of LP production was a few events per minute over a period of about a day², and at Galeras the rate was 1 or 2 events per day over a two-week period²¹. Examples of weak LP swarms were those seen during the waning phase of the 1989–90 Redoubt eruption sequence²⁰. The most energetic LP activity

encountered to date, both in terms of event magnitudes and rate of event production, was that observed during the week preceding the paroxysmal eruption of Mount Pinatubo on 15 June, 1991^{5,50}.

It naturally follows that LP activity may also be triggered by depressurization associated with eruptive activity. This has indeed been observed at Redoubt²⁰ and Galeras³³. At Kilauea, LP activity and weak tremor have been observed following eruptive episodes at Mauna Ulu in 1969, and are often observed during deflation episodes of the Kilauea summit³.

Thus, LP events offer a glimpse into fluid dynamics and can be extremely useful in assessing the internal state of a volcano. To use the forecasting potential of shallow LP events fully, however, one must first be able to distinguish their signatures from those of VT earthquakes, a task made difficult by the extreme heterogeneity of volcanic media.

Source modelling

One tool used by volcano seismologists to discriminate between different source types is the comparison of spectral and temporal

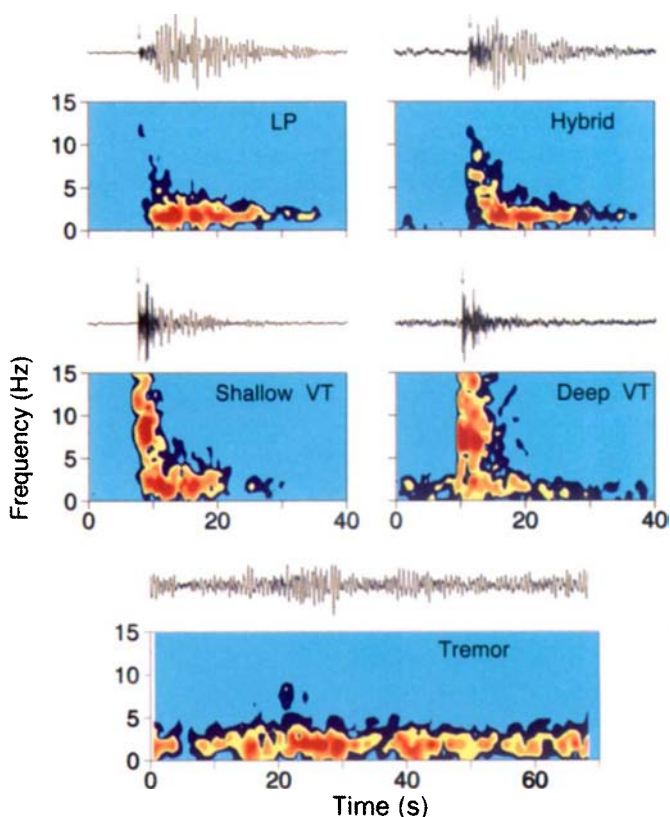


FIG. 2 Vertical ground velocities and associated spectrograms for typical LP, VT and hybrid events and tremor observed at Redoubt volcano. Each spectrogram is a map of the spectral amplitudes of ground velocity versus frequency and time calculated with a moving 1-s window running through the entire seismogram shown above the spectrogram. Warm colours (red, orange and yellow) define the dominant spectral amplitudes; cooler colours (dark and light blue) define lower amplitudes and background. The LP and hybrid events both occurred in a small source zone 1.4 km below the crater floor of Redoubt. The shallow VT earthquake originated 1.7 km below the crater floor, and the deep VT earthquake originated 8.4 km below the crater floor. The tremor displays the same dominant frequency as the LP and hybrid events, but differs from those events in its sustained signal. A small packet of high frequencies 20 s into the tremor seismogram may represent an individual LP event. This tremor is thought to originate from the same source as the LP event. All events were observed at the same receiver location at an epicentral distance of 7.7 km. The origin of the time axis is arbitrary; event onsets are marked by an arrow.

signatures for sources sharing similar locations. Such an approach was taken at Mount St Helens to emphasize the narrow bandwidth of LP events versus the broadband characteristics of VT events³⁰. A similar approach was used at Redoubt (Fig. 2), where the LP signatures were also contrasted to the records of a chemical explosion detonated in the crater, which indicated conclusively that the LP signatures were not due to path or site effects². Another powerful tool is the comparison of the power spectra of LP events recorded at widely separated stations. This technique can be used to show that LP spectra share many common peaks among sites for which the wavefields are uncorrelated, a finding which provides a strong argument in favour of a source effect rather than path or site effects. For shallow sources, resonances in near-surface layers often combine with source resonances to make the identification of spectral peaks in the source more difficult. For repetitive sources, stacking the spectra from many individual events may help the identification task by improving the signal-to-noise ratio, but in other cases there may be no alternative other than to have observations at close range where path effects do not completely mask the distinctive characteristics of LP events. Once common spectral peaks have been identified in the wavefield of a LP event, the next step is to test whether the distribution of such peaks is compatible with the overtones of a resonator. A source model that successfully produces the spectral peaks can then be used to infer source parameters such as source geometry, fluid properties and pressure–space–time history.

Although many geometries are possible resonators, including fluid-filled pipes⁵¹, spheres⁵² and cracks^{38,53,54}, I will focus here on the crack model on the grounds of its ability to explain seismic data while at the same time satisfying mass transport conditions at depth in a volcano. The resonance characteristics of a fluid-filled crack depend on the crack geometry, the physical properties of the fluid and solid, the spatio-temporal characteristics of the pressure fluctuations driving the crack, and the boundary conditions in effect at the crack perimeter. For a rectangular crack the geometrical parameters are expressed by the dimensionless ratios L/d and W/L , where L , W and d represent the length, width and aperture of the crack, respectively. Similarly, the properties of the fluid and solid are expressed by the dimensionless ratios α/a , ρ_s/ρ_f and b/μ , where α is the compressional wavespeed and μ is the rigidity of the solid, a is the acoustic wavespeed and b is the bulk modulus of the fluid, and ρ_s and ρ_f are the densities of the solid and fluid, respectively. Two parameters control the crack resonance, namely the crack stiffness C and impedance contrast Z .

The crack stiffness, defined as^{38,53–55}

$$C = \frac{bL}{\mu d}$$

affects the dispersion characteristics of the crack wave sustaining the crack resonance and thereby fixes the resonant frequencies produced by the source. The phase velocity of the crack wave decreases with increasing values of the crack stiffness, so that a crack with large aspect ratio L/d or large contrast b/μ may be capable of producing long-period signals. The impedance contrast between fluid and solid, defined as^{38,55}

$$Z = \frac{\rho_s \alpha}{\rho_f a}$$

affects both the frequency and duration of the radiated signal. The presence of gas bubbles can drastically reduce the sound speed of the fluid⁵⁶ so that resonance at a long period may be possible in a short crack. The effect of these bubbles is also to increase the impedance contrast, which increases the duration of the signal. Fluid viscosity contributes another parameter to the description of the source that describes viscous energy loss in addition to radiation loss. Besides its direct effect on signal duration, this viscous dissipation factor does not contribute much to source complexity^{38,54} and is not discussed further here.

The spectrum radiated by the crack represents the interference pattern between the longitudinal and lateral modes of resonance, which depends on the ratio W/L and the two parameters C and Z . Beyond these factors, which represent the coupling between the elastodynamics of the solid and acoustics of the fluid via the crack geometry, source complexity is a direct result of the spatio-temporal characteristics of the pressure field associated with the rich nonlinear dynamics of the fluid.

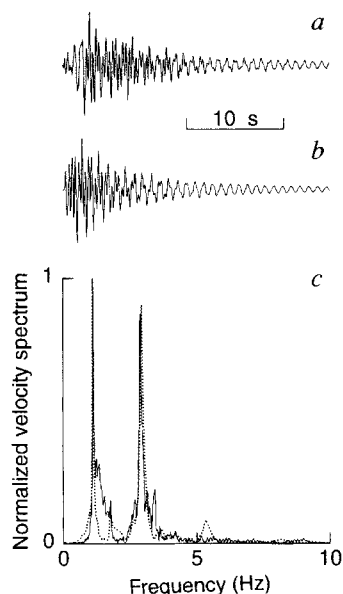
Long-period events recorded at Galeras (Fig. 1) offer graphic examples of the effect of impedance contrast on the duration of the signal. For example, Fig. 3 compares the Galeras data to synthetic seismograms calculated for a vertical fluid-filled crack of rectangular shape embedded in a homogeneous half-space. The synthetics represent the ground response to the crack excitation owing to a sudden step in pressure applied over a small area of the crack wall near the centre of the crack. The good match between data and synthetics suggests that, to first order, the slot-like fluid-filled crack model provides a simple geometry consistent with observations. The fit requires an impedance contrast of 15, which is strongly suggestive of a gas-filled cavity.

The situation at Galeras may have been close to optimum in terms of the relative importance of source versus path effects. The source depth of $1(\pm 0.2)$ km and epicentral distance of $1(\pm 0.1)$ km were such that multiple reflections in thin soft surficial layers draping the volcano were apparently minimized. At the same time the sustained resonance of the source associated with the strong impedance contrast sharpened the definition of the spectral peaks (Fig. 3c) making their identification easier. The situation is considerably more complex when the source is embedded in soft shallow layers and the impedance contrast is weak.

To illustrate the effect of medium heterogeneity on the radiation from a low-impedance fluid-driven crack, I now consider the case of a Hawaiian basaltic magma. Figure 4a compares synthetic seismograms of the vertical component of ground velocity calculated for a vertical crack embedded in a layered structure with those calculated for the same source embedded in a homogeneous half-space. The fluid in the crack has a sound speed of 1.25 km s^{-1} , which is compatible with a basaltic magma containing a void fraction of gas of about 0.05%, and the impedance contrast is near 1. The layered structure is appropriate for the area of Pu'u O'o crater on the east rift of Kilauea⁵⁷ (see Fig. 4 legend), and the homogeneous half-space has parameters identical to those found in layer 3 of that structure.

The low impedance contrast accounts for the poorly developed source resonance, as demonstrated by the weak harmonic coda in the synthetics calculated for the homogeneous medium. Whereas

FIG. 3 Comparison between data and synthetics for an LP event that preceded the 14 January 1993 eruption of Galeras volcano. a, Vertical ground velocity recorded at a distance of 1 km from the crater. b, Vertical ground velocity calculated at the same location for the excitation of a vertical fluid-filled crack buried at a depth of 1 km beneath the vent of Galeras. c, Spectrum for data (thin line) and synthetics (dotted line). The source model has the following parameters: crack length, 180 m; crack width, 90 m; crack aperture, 0.05 m; crack stiffness, 100; sound speed of fluid, 0.3 km s^{-1} ; compressional wavespeed of solid, 2.5 km s^{-1} ; density ratio of fluid to rock, ~ 0.5 ; and ratio of bulk modulus of fluid to rigidity of rock, 0.025.



these synthetics are dominated by sharp P, S and Rayleigh phases, those derived for the layered medium present a distinctive spindle-shaped signal, in which crack oscillations appear as a short-duration quasi-monochromatic coda emerging out of the broader-band contributions from near-surface layer resonances. At first glance, this would seem to make the task of distinguishing

the LP signature from VT events rather hopeless. Fortunately, the spectral character of the signal is helpful here.

The spectra derived from the synthetics (Fig. 4b) show that the source has distinct peaks which can be identified among spectral peaks due to path effects. For example, source effects are clearly identified in the homogeneous half-space where all spectra show a peak near 3.8 Hz due to crack resonance (see arrows above individual spectra). Other peaks due to the source are also present, although less obvious because of the dependence of their excitation on the position of the pressure transient triggering the crack resonance. The dominant peak near 3.8 Hz is also present in the spectra obtained for the layered structure, the effect of which is to introduce additional peaks due to near-surface resonances. The peaks associated with path effects, however, are strongly dependent on the source–receiver configuration, whereas those due to the source are not. This is demonstrated in Fig. 5a for a crack buried at various depths in the Pu'u O'o structure.

The synthetics shown in Fig. 5a incorporate the same source–receiver geometry and model parameters but assume the position of the pressure transient to be fixed at the centre of the crack. Three source depths ($h = 0, 105$ and 300 m) are used to assess the relative importance of path effects for sources that reach up to the surface, reach up to the bottom of layer 2, or are completely decoupled from those shallow layers. Both homogeneous and layered media produce roughly the same spectra, which all display a peak at a frequency near 3.8 Hz due to the source. The effect of structure on the spectra of the deeper sources in or at the top of the half-space (layer 3 in the layered model) is insignificant (see bottom two panels for $h = 105$ and 300 m in Fig. 5a), and although path effects are stronger for the surficial source (see top panel in Fig. 5a), the dominant source peak is still easily identifiable. These calculations demonstrate that, to first order, source resonances are stable features in the radiated spectra, and suggest that these resonances may be separated from resonances due to path effects even for low-impedance sources. In practical situations, the identification of source effects can be enhanced through the expedient of stacking data from distinct receivers.

The above model may be used to illustrate the relationship between LP events and sustained tremor. Two examples in Fig. 5b compare models of tremor generation in the Pu'u O'o structure. The model geometry is identical to that used in Fig. 5a. I assumed that a complex pressure–time history could cause sequential excitation of the crack and modelled this source excitation by superposing identical waveforms with random time delays. The original waveforms generated by the step excitation of the crack and the waveforms obtained by superposition of delayed versions of those synthetics are shown at the top of the figure for two crack depths. An identical sequence of random time delays between 0 and 1 s generated from five-digit random numbers⁵⁸ was used to produce the latter results. The spectra of the resulting synthetics (see bottom two panels in Fig. 5b) share common peaks although they differ markedly in their detailed structure. The spectra shown in the bottom panel, in particular, suggest that source excitation complexity can significantly affect the radiated spectrum by splitting and modulating the amplitudes of the resonant peaks associated with the step response and by contributing additional peaks to that spectrum. Of particular interest is the strong dependence of the band of radiated energy on source depth, due in large part to the amplification and splitting of spectral peaks below 4 Hz in the deeper source (bottom panel) produced by random excitations. These waveforms and spectra agree qualitatively with those observed for tremor at Pu'u O'o⁵⁹ and indicate that path effects and effects that are characteristic of the fluid dynamics may both contribute to the observed spectral complexity for shallow sources. Obviously, the complexities of the driving mechanisms of tremor make this a much more difficult source to study than the LP event, and it is only under the more favourable circumstances of a tremor source with stationary spatio-temporal characteristics that quantification of the source mechanisms may be feasible.

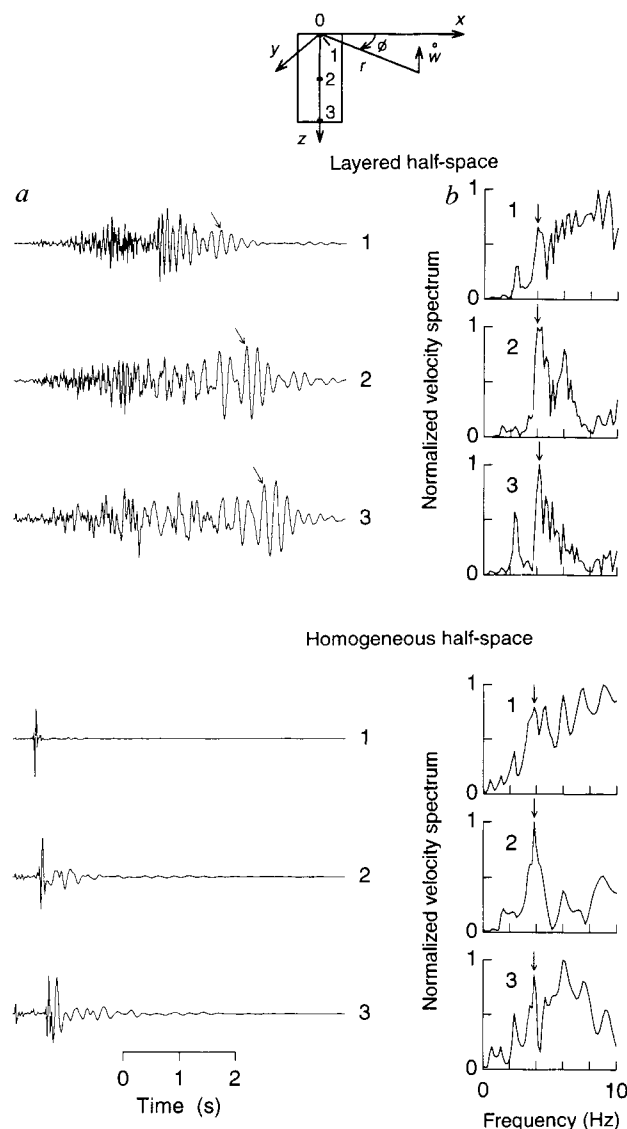


FIG. 4 Comparisons between ground velocity responses due to the excitation of a fluid-filled crack embedded in a layered half-space and those obtained for the same source buried in a homogeneous half-space. The layered half-space has three layers, with the following P- and S-wave velocities (km s^{-1}), densities (g cm^{-3}) and thicknesses (m): layer 1 (0.65, 0.35, 2.65, 30); layer 2 (1.25, 0.72, 2.70, 75); layer 3 (2.50, 1.44, 2.75, ∞). The homogeneous half-space has parameters identical to those of layer 3 in that structure. The crack geometry is shown in the Cartesian coordinates at the top of the figure. All the synthetics are for a 100-m-wide vertical crack extending from the free surface to a depth of 200 m. The crack is excited by a step in pressure applied over a small area of the crack just below the free surface, at 100 m depth, or at 200 m depth. Each synthetic represents a vertical component $\dot{w}$ calculated at a distance r of 1 km and azimuth ϕ of 25 degrees. The numbers (1, 2, 3) attached to the seismograms and spectra refer to the locations of the pressure transient shown in the sketch. a, Synthetic seismograms. The slanted arrows point to the short-duration quasi-monochromatic coda associated with crack resonance in the synthetics calculated for the layered structure. b, Spectra of the seismograms shown at the left. The arrows mark one of the resonant peaks due to the source; this spectral peak is present in all models.

Under such circumstances the technique of spectral stacking may again be helpful in extracting stationary peaks due to sustained source resonance.

Quantitative analysis at Redoubt volcano

The reawakening of Redoubt volcano on 13 December, 1989, after 23 years of quiescence, was heralded by the gradual onset of repetitious LP events. This activity soon evolved into an intense swarm that lasted 23 hours and terminated with the initial major eruption on 14 December, which blew ash to an altitude in excess of 10,000 m (refs 2,60). Several outstanding features of the swarm were immediately recognized², as follows: a rapidly accelerating rate of seismic energy release over the first 18 hours of activity, followed by a decline coinciding with the emergence of sustained tremor in the final 5 hours before the eruption; nearly identical signatures for individual events in the swarm; dilatational first motions on all stations of the network for all LP events with clear P phases; and a magnitude range from -0.4 to 1.6 for the entire sequence. Detailed spectral analyses performed subsequently revealed that both LP events and tremor had similar spectra

marked by several dominant and subdominant peaks common to stations covering a wide range of azimuths and distances from the source², while careful location and error analyses established that this activity originated from a single point source 1.4 km below the crater floor³⁶.

These features were found to be compatible with those produced by the resonant excitation of a fluid-driven crack and yielded the following parameters for the LP source²: crack length, 280–380 m; crack width, 140–190 m; crack aperture, 0.05–0.20 m; crack stiffness, 100–200; sound speed of fluid, 0.8–1.3 km s⁻¹; compressional wavespeed of rock, 5.1 km s⁻¹; density ratio of fluid to rock, about 0.4; and ratio of bulk modulus of fluid to rigidity of rock, 0.03–0.07. Crack excitation was fitted by an impulsive pressure drop in the range 0.4–40 bar. These parameters, along with the spatio-temporal characteristics of the pressure transient and observed intermittency, collectively were found to be consistent with forced crack excitations associated with unsteady choked flow conditions².

A conceptual model of the processes that may have characterized the pre-eruptive sequence observed at Redoubt is shown in Fig. 6a. The sketch shows the old dome capping the vent, the hydrothermal system, shallow magma chamber, and LP source located between these two reservoirs. The evidence for a shallow magma chamber comes from a tomographic study conducted at Redoubt in 1991, in which a volume of approximately 2–3 km³ with P- to S-wave velocity ratio 4% higher than average was found centred 2.5 km beneath the crater and 1–2 km north of the eruptive vent⁶¹. This small volume is located immediately beneath and to the north of the source of the LP swarm. Petrological data provide evidence for a magma-mixing event preceding the start of the eruption⁶², which is thought to have been the source of the excess pressure driving the sequence⁶³. Accordingly, the initiation of the swarm may be interpreted as representing the failure of a pressure-relief valve connecting a deeper supercharged magma-dominated reservoir to a shallow low-pressure hydrothermal system. The sustained rate of 3–5 LP events per minute observed during the first 18 hours of the swarm² may then be viewed as the result of unsteady choking of a supersonic flow of magmatic steam driven by the pressure gradient existing between the two reservoirs. By the same token, the emergence of sustained tremor 5 hours before the eruption may represent a change in choked-flow regime associated with a gradual weakening of the pressure gradient driving the flow⁶⁴.

The graph in Fig. 6b gives a schematic representation of the pressure histories in the magma chamber and hydrothermal system, and pressure gradient between these two reservoirs, along with significant events inferred to have occurred during the 23-hour-long pre-eruptive phase. Injection of a more mafic magma into the base of the magma reservoir⁶² represents the initial event that triggered magmatic convection and increased exsolution of volatiles leading to rapid pressurization of the system. At some time during this pressurization phase, critical pressure was reached and either a fracture was created or a pre-existing fissure was dilated and communication was opened to the overlying lower-pressure hydrothermal system. The pressure gradient between the two systems then drove magmatic steam through this conduit, and LP activity was triggered by unsteady choked flow in response to fluctuations of outlet pressure associated with the reaction of the upper reservoir to the injection of mass and heat from below⁶⁵. Initially, the hydrothermal system may have responded to the increased heat injection into its base by accelerated convection that would have manifested itself in increased steam emissions at the summit of Redoubt, and indeed observations of the volcano a few days before the eruption seem to indicate that such may have been the case. Later in the swarm sequence, the import of heat started to exceed the capacity of the hydrothermal system to evacuate so that heat and pressure started to rise in that system as well. Eventually, the pressure gradient driving the LP source started to decrease and the flow regime inside the crack changed from intermittent, high-

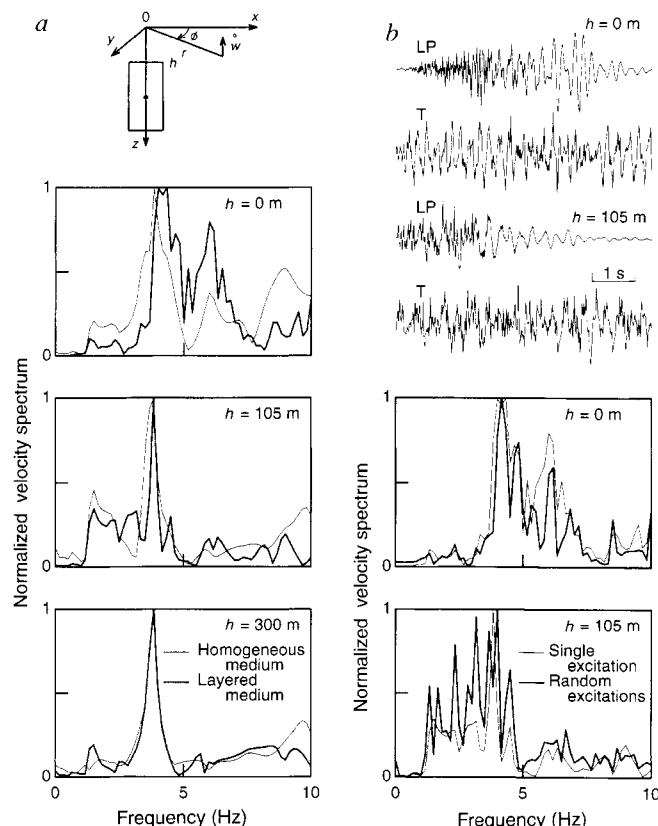
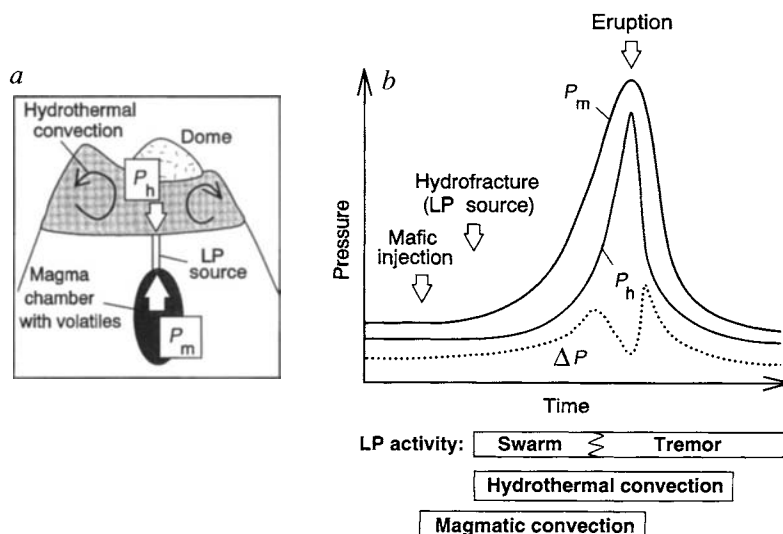


FIG. 5 Dependence of synthetics on crack depth and crack excitation. The crack geometry is shown in the cartesian coordinates at the upper left. All synthetics are for a 100-m-wide, 200-m-long, vertical crack excited by a pressure transient applied at the centre of the crack. Each synthetic represents a vertical component of ground velocity w calculated at a distance r of 1 km and azimuth ϕ of 25° . a, Comparisons between spectra obtained for a crack buried in a homogeneous half-space and those derived for the same crack embedded in a layered half-space. The spectra are calculated for a step excitation of a crack buried at depths $h = 0$ m (top), $h = 105$ m (middle), and $h = 300$ m (bottom). The spectra shown for $h = 0$ m are the same as those depicted for case 2 in Fig. 4b. b, Comparisons between ground velocities and spectra calculated for a single step excitation and those derived for multiple step excitations of a crack embedded in a layered structure. Top, synthetics obtained for a single excitation (LP) and synthetics obtained by delaying and superposing multiple copies of the LP waveforms (T) for two crack depths $h = 0$ m and $h = 105$ m. Bottom, spectra of the seismograms shown above.

FIG. 6 Conceptual model of pre-eruptive processes at Redoubt volcano on 13–14 December 1989. *a*, Sketch illustrating the spatial relationship between the stationary LP source, the magma chamber inferred from tomography, the hydrothermal system, and lava dome emplaced during the previous eruption in 1966. *b*, Cartoon sketching the pressurization of the magma chamber, P_m , and associated pressurization of the hydrothermal system, P_h , along with the pressure difference between these two systems, ΔP . Significant events in the pressurization history are indicated by arrows. Note that the pressure histories depicted are heavily smoothed to enhance gross behaviour (in reality, these curves are expected to display small-scale fluctuations superimposed on the smoothed trends depicted). Episodes of LP activity, hydrothermal convection and magmatic convection are shown in boxes beneath the graph, where the horizontal extent of the boxes represents the observed duration of LP activity and inferred durations of hydrothermal and magmatic convections.



amplitude pressure transients producing discrete LP events, to sustained, low-amplitude pressure fluctuations producing tremor. The pressure gradient between the two reservoirs continued to decrease for a few hours as the pressure kept on rising in both of them, finally culminating in the dome-destroying eruption.

Future directions

The above considerations point to the importance of shallow LP activity in forecasting the onset and climactic stages of an important class of highly hazardous volcanic eruptions—namely, the stratovolcanoes bordering the highly populated western margin of the Americas, as well as those of the Mexican–Caribbean, Mediterranean and Pacific Basin regions of the world. To benefit fully from this new understanding of the processes underlying volcano behaviour, however, further elaboration of the physical model of LP activity will be required. Foremost among the challenges facing volcano seismologists is a better understanding of the excitation mechanisms of LP events and tremor in relation to the overall process of fluid percolation in a volcano⁴³. The challenge is best exemplified in the activity of deep LP events (> 10 km depth).

Although a direct link has been observed between shallow LP activity and eruptions, the relationship with deeper LP events is less clear. Long-period events occurred at depths near 30 km at Pinatubo in early June 1991, where they were attributed to a basaltic intrusion which is thought to have triggered the processes leading to the cataclysmic 15 June eruption^{50,66}. On the other hand, LP events are regularly observed at depths of 30–60 km beneath Kilauea, where their occurrence appears to be more directly related to deep magma supply dynamics than surface activity⁴³. A low background of deep LP events is not uncommon, and may indeed be the rule rather than the exception as accumulating observations now increasingly suggest^{67–70}. The difficulty here has to do with our lack of knowledge concerning the character and dynamics of deep-seated fluid transport under volcanoes. Another difficulty stems from the wide range of time-scales involved in volcanic processes. Short-period seismology, which typically covers the bandwidth 1–20 Hz, is not by itself sufficient to resolve longer-term processes such as those associated with mass transport. For example, the timescale of mass transport between the magma and hydrothermal reservoirs during the precursory sequence leading to the 1989 eruption of Redoubt volcano is about a day, which is well beyond the response capability of short-period instruments. To document fully the relationship between point-like LP activity and the overall pathway structure used by eruptive fluids, broadband instruments are required. Sacks–Evertson borehole strain-meters have proven to be quite useful in volcanic applications⁷¹, and arrays of such instruments coupled with three-component broadband seismo-

graphs may present unprecedented opportunities to monitor the movements of magmatic and hydrothermal fluids before eruptions. Progress in broadband volcano seismology has already been made in that direction with encouraging results^{72–74}.

A fuller understanding of the excitation mechanisms of LP events and tremor may be achieved through further modelling efforts. Kinematic models such as those discussed above (in which conduit resonance is excited via an arbitrary initial condition; see Figs 3–5) will need to be replaced by dynamic models (in which the conduit is self-excited by some disturbance in the flow process) if we are to understand better the link between seismic signals and fluid processes in a volcano. New studies have brought to light self-excitation mechanisms inherent to fluid nonlinearity that point the way in that direction^{65,75}. With the advent of more powerful computers, numerical studies of multi-phase flows are improving our knowledge of both the micro- and macrophysics of such flows. Computer models can now be developed to study various types of flow instabilities⁷⁶, and laboratory experiments can be used to anchor numerical results and give further insights into self-excited pressure oscillations^{64,75} that may be adequate models for LP and tremor sources. □

B. A. Chouet is at the US Geological Survey, 345 Middlefield Road, MS 977, Menlo Park, California 94025, USA.

- Shaw, H. R. *J. geophys. Res.* **90**, 11275–11288 (1985).
- Chouet, B. A., Page, R. A., Stephens, C. D., Lahr, J. C. & Power, J. A. *J. Volcan. geotherm. Res.* **62**, 95–135 (1994).
- Koyanagi, R. Y., Chouet, B. & Aki, K. *Prof. Pap. U.S. geol. Surv.* **1350**, 1221–1257 (1987).
- Endo, E. T., Malone, S. D., Noson, L. L. & Weaver, C. S. *Prof. Pap. U.S. geol. Surv.* **1250**, 93–107 (1981).
- Pinatubo Volcano Observatory Team *Eos* **72**, 545–555 (1991).
- Power, J. A., Jolly, A. D., Page, R. A. & McNutt, S. R. *Bull. U.S. geol. Surv.* **2139**, 149–159 (1995).
- Tilling, R. I. & Dvorak, J. *Nature* **363**, 125–133 (1993).
- Lockwood, J. P. et al. *Eos* **66**, 169–171 (1985).
- Ida, Y. *Nature* **352**, 571–572 (1991).
- Venzke, E., Wunderman, R. & Peng, G. (eds) *Smithsonian Instn Bull. Global Volc. Net.* **19**, 4–7 (1994).
- Newhall, C. G. & Dzurisin, D. *Bull. U.S. geol. Surv.* **1855**, 7–16 (1988).
- McNutt, S. R. *Encyclopedia of Earth Science* Vol. 4, 417–425 (Academic, 1992).
- Barberi, F., Bertagnini, A., Landi, P. & Principe, C. *J. Volcan. geotherm. Res.* **52**, 231–246 (1992).
- Minakami, T. in *Physical Volcanology* (eds Civetta, L., Gasparini, P., Luongo, G. & Rapolla, A.) 1–27 (Elsevier, New York, 1974).
- Kagiyama, T., Gyoza, N., Koyama, E. & Tsuji, H. in *Comparative Studies of Physical Background of Volcanic Activity and its Relation to Eruption Disasters* (ed. Okada, H.) 92–101 (Hokkaido Univ., Hokkaido, 1985).
- Shimozuru, D. & Kagiyama, T. in *Volcanic Hazards* (ed. Latter, J. H.) 504–512 (Springer, Berlin, 1989).
- Rep. Coordin. Committee on Prediction of Volcanic Eruptions Vol 41, 64–77 (1988).
- Ukawa, M. *J. Volcan. geotherm. Res.* **55**, 33–50 (1993).
- Havskov, J., De la Cruz-Reyna, S., Singh, S. K., Medina, F. & Gutiérrez, C. *Geophys. Res. Lett.* **10**, 293–296 (1983).
- Stephens, C. D., Chouet, B. A., Page, R. A., Lahr, J. C. & Power, J. A. *J. Volcan. geotherm. Res.* **62**, 153–182 (1994).
- Muñoz, F. A. et al. *Eos* **74**, 281–287 (1993).

22. Okada, H., Nishimura, Y., Miyamachi, H., Mori, H. & Ishihara, K. *Bull. volcan. Soc. Jap.* **35**, 175–203 (1990).
23. Gil Cruz, F., Meyer, H. J., Chouet, B. & Harlow, D. *Hawaii Symp. on How Volcanoes Work, Diamond Jubilee (1912–1987)* 90 (Hawaiian Volcano Observatory, Hilo, Hawaii, 1987).
24. Martinelli, B. J. *Volcan. geotherm. Res.* **41**, 297–314 (1990).
25. Venzke, E., Wunderman, R. & Peng, G. (eds) *Smithsonian Instn Bull. Global Volc. Net.* **17(6)**, 11 (1992).
26. Malone, S. D., Endo, E. T., Weaver, C. S. & Ramey, J. W. *Prof. Pap. U.S. geol. Surv.* **1250**, 803–813 (1981).
27. Qamar, A., St Lawrence, W., Moore, J. N. & Kendrick, G. *Bull. seism. Soc. Am.* **73**, 1797–1813 (1983).
28. Swanson, D. A. et al. *J. Geodynamic.* **3**, 397–423 (1985).
29. Weaver, C. S., Grant, W. C., Malone, S. D. & Endo, E. T. *Prof. Pap. U.S. geol. Surv.* **1250**, 109–121 (1981).
30. Fehler, M. & Chouet, B. *Geophys. Res. Lett.* **9**, 1017–1020 (1982).
31. Fischer, T. P. et al. *Nature* **368**, 135–137 (1994).
32. Nishi, K. A. *Disaster Prev. Res. Inst. Kyoto Univ.* **27 B-1**, 29–34 (1984).
33. Venzke, E., Wunderman, R. & Peng, G. (eds) *Smithsonian Instn Bull. Global Volc. Net.* **17(12)**, 2–4 (1992).
34. Venzke, E., Wunderman, R. & Peng, G. (eds) *Smithsonian Instn Bull. Global Volc. Net.* **18(6)**, 2–3 (1993).
35. Minakami, T. *Bull. Earthquake Res. Inst. Univ. Tokyo* **38**, 497–544 (1960).
36. Lahr, J. C., Chouet, B. A., Stephens, C. D., Power, J. A. & Page, R. A. *J. Volcan. geotherm. Res.* **62**, 137–151 (1994).
37. Eaton, J. P., Richter, D. H. & Krivoy, H. L. *Prof. Pap. U.S. geol. Surv.* **1350**, 1307–1336 (1987).
38. Chouet, B. in *Volcanic Seismology* (eds Gasparini, P., Scarpa, R. & Aki, K.) 133–156 (Springer, Berlin, 1992).
39. Shaw, H. R. in *Physics of Magmatic Processes* (ed. Hargraves, R. B.) 201–264 (Princeton Univ. Press, 1980).
40. Takeo, M. *Phys. Earth planet. Inter.* **32**, 241–264 (1983).
41. Peck, D. L. & Minakami, T. *Geol. Soc. Am. Bull.* **79**, 1151–1166 (1968).
42. Chouet, B. *J. geophys. Res.* **84**, 2315–2330 (1979).
43. Shaw, H. R. & Chouet, B. *J. geophys. Res.* **96**, 10191–10207 (1991).
44. Hill, D. P. *J. geophys. Res.* **82**, 1347–1352 (1977).
45. Aki, K. & Koyanagi, R. *J. geophys. Res.* **86**, 7095–7109 (1981).
46. Kamo, K., Furuzawa, T. & Akamatsu, J. *Bull. Volcan. Soc. Japan.* **22**, 41–58 (1977).
47. Wolfe, E. W. *Earthquakes & Volcanoes* **23(1)** 5–37 (1992).
48. Ukawa, M. *J. Volcan. geotherm. Res.* **55**, 33–50 (1993).
49. Miller, T. P. *J. Volcan. geotherm. Res.* **62**, 197–212 (1994).
50. White, R. A., Harlow, D. H. & Chouet, B. A. *Eos (suppl.)* **73(43)**, 347 (1992).
51. Chouet, B. *J. geophys. Res.* **90**, 1881–1893 (1985).
52. Crosson, R. S. & Bame, D. A. *J. geophys. Res.* **90**, 10237–10247 (1985).
53. Chouet, B. *J. geophys. Res.* **91**, 13967–13992 (1986).
54. Chouet, B. *J. geophys. Res.* **93**, 4373–4400 (1988).
55. Aki, K., Fehler, M. & Das, S. J. *Volcan. geotherm. Res.* **2**, 259–287 (1977).
56. Kieffer, S. W. *J. geophys. Res.* **82**, 2895–2904 (1977).
57. Ferrazzini, V., Aki, K. & Chouet, B. *J. geophys. Res.* **96**, 6199–6209 (1991).
58. Abramowitz, M. & Segun, I. A. *Handbook of Mathematical functions* 991 (Dover, New York, 1972).
59. Goldstein, P. & Chouet, B. *J. geophys. Res.* **99**, 2637–2652 (1994).
60. Brantley, S. R. *U.S. geol. Surv. Circ.* **1061**, 1–33 (1990).
61. Benz, H. M. et al. *J. Geophys. Res.* (in the press).
62. Swanson, S. E., Nye, C. J., Miller, T. P. & Avery, V. F. *J. Volcan. geotherm. Res.* **62**, 453–468 (1994).
63. Sparks, S. R. J., Sigurdsson, H. & Wilson, L. *Nature* **267**, 315–318 (1977).
64. Meier, G. E. A., Grabitz, G., Jungowsky, W. M., Wiltczak, K. J. & Anderson, J. S. *Mitteilungen aus dem Maz-Plank Institut für Strömungsforschung und der Aerodynamischen Versuchsanstalt* **65**, 1–172 (1978).
65. Morrissey, M. M. & Chouet, B. A. (abstr.) *IUGG XXI General Assembly A461* (IUGG, Boulder, 1995).
66. Pallister, J. S., Hoblitt, R. P. & Reyes, A. G. *Nature* **356**, 426–428 (1992).
67. Pitt, A. M. & Hill, D. P. *J. geophys. Res. Lett.* **21**, 1679–1682 (1994).
68. Ukawa, M. *Nat. Res. Center for Disaster Prev.* **37**, 129–133 (1984).
69. Ukawa, M. & Ohtake, M. *J. geophys. Res.* **92**, 12649–12663 (1987).
70. Hasegawa, A., Zhao, D., Hori, S., Yamamoto, A. & Honuchi, S. *Nature* **352**, 683–689 (1991).
71. Linde, A. T., Agustsson, K., Sacks, I. S. & Stefansson, R. *Nature* **365**, 737–740 (1993).
72. Kanamori, H., Given, J. W. & Lay, T. *J. geophys. Res.* **89**, 1856–1866 (1984).
73. Uehira, K. & Takeo, M. *J. geophys. Res.* **99**, 17775–17789 (1994).
74. Kawakatsu, H., Ohminato, T. & Ito, H. *Geophys. Res. Lett.* **21**, 1963–1966 (1994).
75. Julian, B. R. *J. geophys. Res.* **99**, 11859–11877 (1994).
76. Harlow, F. H. & Amsden, A. A. *J. comput. Phys.* **17**, 19–52 (1975).
77. Meier, G. E. A., Szumowski, A. P. & Selerowicz, W. C. *Prog. Aerospace Sci.* **27**, 145–200 (1990).

ACKNOWLEDGEMENTS. I thank H. R. Shaw and R. A. Page for discussions and comments on the manuscript; C. D. Stephens and J. C. Lahr for help in developing the pressurization model of Redoubt volcano; C. D. Stephens and P. B. Dawson for their help in preparing illustrations; and D. P. Hill, J. A. Power and M. Iyer for reviews.

Structural similarity between TAFs and the heterotetrameric core of the histone octamer

Xiaoling Xie^{*†}, Tetsuro Kokubo^{‡#}, Steven L. Cohen[§], Urooj A. Mirza[§], Alexander Hoffmann^{||#}, Brian T. Chait[§], Robert G. Roeder^{||}, Yoshihiro Nakatani[‡] & Stephen K. Burley^{*†||}

Laboratories of ^{*} Molecular Biophysics, [§] Mass Spectrometry and Gaseous Ion Chemistry, and ^{||} Biochemistry and Molecular Biology, and [†] Howard Hughes Medical Institute, The Rockefeller University, 1230 York Avenue, New York, New York 10021, USA
[‡] National Institutes of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland 20892, USA

A complex of two TFIID TATA box-binding protein-associated factors (TAF_{II}s) is described at 2.0 Å resolution. The amino-terminal portions of dTAF_{II}42 and dTAF_{II}62 from *Drosophila* adopt the canonical histone fold, consisting of two short α -helices flanking a long central α -helix. Like histones H3 and H4, dTAF_{II}42 and dTAF_{II}62 form an intimate heterodimer by extensive hydrophobic contacts between the paired molecules. In solution and in the crystalline state, the dTAF_{II}42/dTAF_{II}62 complex exists as a heterotetramer, resembling the (H3/H4)₂ heterotetrameric core of the histone octamer, suggesting that TFIID contains a histone octamer-like substructure.

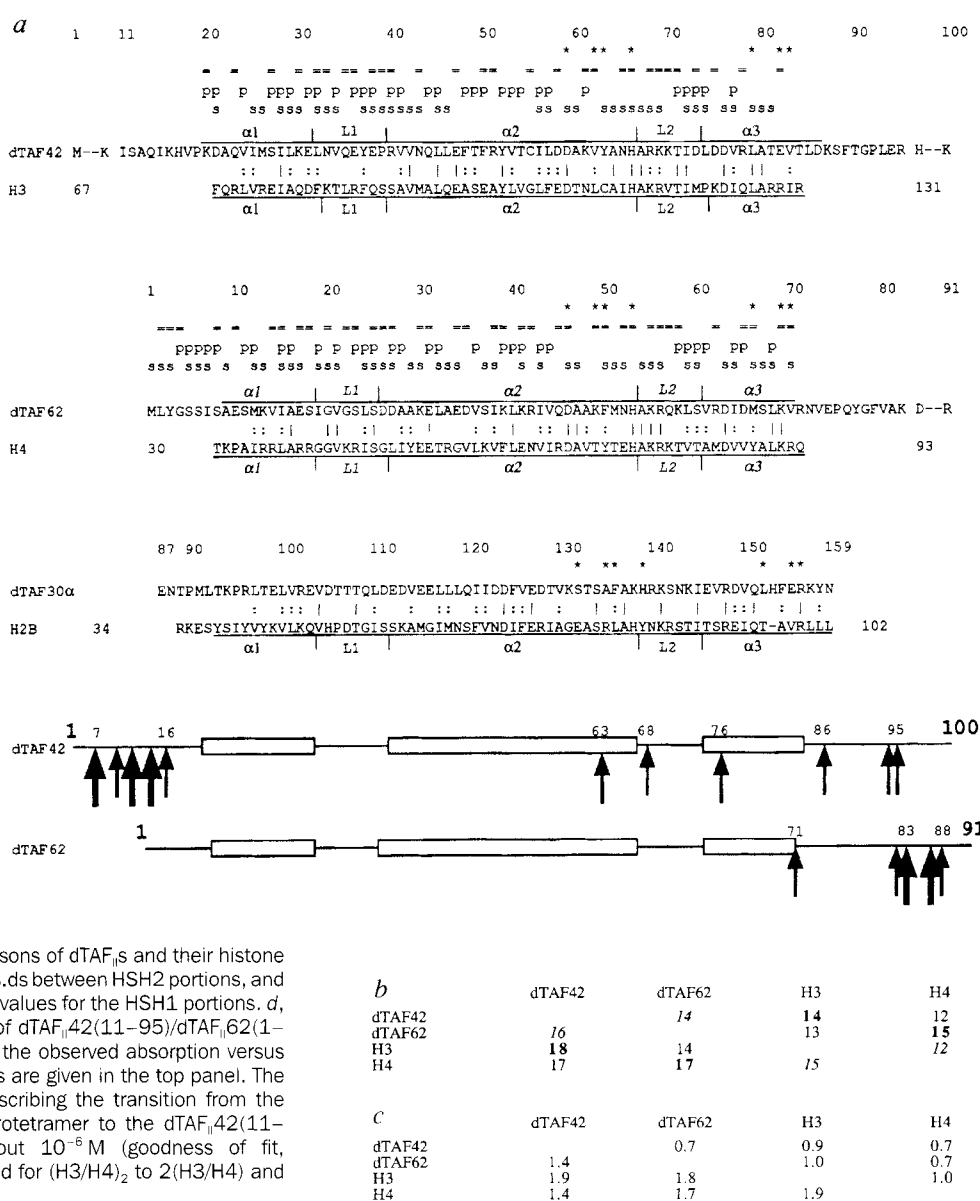
EUKARYOTIC transcription initiation and its regulation are best understood for genes transcribed by RNA polymerase II (Pol II) with the general transcription factors TFIIA, TFIIB, TFIID, TFIIE, TFIIIF, TFIIIG/J, TFIH and TFIIH (reviewed in refs 1,2). In some cases, this process begins with recognition of the TATA element within the core promoter by the DNA-binding subunit of

TFIID (TATA box-binding protein, TBP), forming a multi-protein–DNA complex that coordinates accretion of class II initiation factors and Pol II into a preinitiation complex (PIC) (reviewed in ref. 3). TFIIB is the next general factor to enter the PIC, creating a TFIIB–TFIID–DNA platform that is recognized by Pol II plus TFIIIF. *In vivo*, Pol II transcription depends on TFIIE and TFIH, and possibly TFIIA. Once PIC assembly is complete, and nucleoside triphosphates are present, strand separation occurs to give an open complex, the large subunit of Pol II is phosphorylated, and Pol II initiates transcription and is

^{*}To whom correspondence should be addressed.

[#] Present address: Division of Gene Function in Animals, Nara Institute of Science and Technology, 8916-5 Takayama, Ikoma, Nara 630-01, Japan (T.K.); Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA (A.H.).

FIG. 1 a, Top, sequence alignments of histone-like portions of dTAF₃₀ α , dTAF₄₂ and dTAF₆₂ with chicken H2B, H3 and H4, respectively. Amino-acid identities are denoted by vertical lines and similarities with colons. The α -helical regions were assigned from X-ray structures³⁴. Residues are labelled with 's', 'p' and an asterisk, indicating involvement in intra-molecular, heterodimer, and heterotetramer contacts, respectively. For H2B, H4, dTAF₃₀ α and dTAF₆₂, the asterisk denotes residues involved in contacts between distinct heterodimers. Solvent-accessible residues in the (dTAF₄₂(17–86))/(dTAF₆₂(1–70))₂ heterotetramer are denoted by short black bars. Bottom, proteolysis of the dTAF₄₂(1–100)/dTAF₆₂(1–91) complex. Endoproteases Asp-N, Glu-C, trypsin, subtilisin and chymotrypsin were used as previously described²⁹. The polypeptide chains are represented schematically with large arrows denoting cleavage sites observed within minutes to hours, and small arrows denoting cleavage sites observed between 4 and 24 hours. b, Sequence comparisons of dTAF₄₂(22–83), dTAF₆₂(9–70), H3(68–130) and H4(31–92). The upper triangle gives the number of identical amino acids, and the lower triangle gives the number of chemically similar amino acids. Bold indicates comparisons between homologous proteins, and italics denote comparisons between heterodimerization partners. c, Pairwise comparisons of dTAF_is and their histone homologues. The upper triangle gives r.m.s.ds between HSH2 portions, and the lower triangle gives the corresponding values for the HSH1 portions. d, Equilibrium analytical ultracentrifugation of dTAF₄₂(11–95)/dTAF₆₂(1–82). The fitted curve is superimposed on the observed absorption versus radius profile, and the calculated residuals are given in the top panel. The equilibrium dissociation constant (K_d) describing the transition from the (dTAF₄₂(11–95)/dTAF₆₂(1–82))₂ heterotetramer to the dTAF₄₂(11–95)/dTAF₆₂(1–82) heterodimer is about 10^{-6} M (goodness of fit, 0.00012). Similar K_d values were obtained for (H3/H4)₂ to 2(H3/H4) and its reverse reaction³⁵. METHODS. *Drosophila* TAF₄₂ residues 11–95 and *Drosophila* TAF₆₂ residues 1–82 fused with GST were separately overexpressed as inclusion bodies in *E. coli*. Denatured dTAF₄₂(11–95) and GST-dTAF₆₂(1–82) were mixed in an equimolar ratio, co-renatured by stepwise dialysis, treated with protease to remove the GST tag, and purified to homogeneity by conventional methods. Mass spectrometry showed that the dTAF₄₂(11–95)/dTAF₆₂(1–82) complex used successfully for crystallization was neither modified nor further proteolysed during expression, renaturation and purification. The measured relative molecular mass values were: dTAF₄₂(11–95), 9.917 ± 1 (predicted, 9,918); and dTAF₆₂(1–82), 9.071 ± 1 (predicted, 9,072). Dynamic light scattering was performed with a Molecular Size Detector (Protein Solutions, Charlottesville, VA). Equilibrium analytical ultracentrifugation was performed using a Beckman model XL-A ultracentrifuge.



released from the promoter. During elongation *in vitro*, TFIID can remain bound to the core promoter and support rapid reinitiation of transcription (reviewed in ref. 4). An abbreviated PIC assembly mechanism has also been proposed, following discoveries of various Pol II holoenzymes (reviewed in ref. 5).

The role of TFIID in eukaryotic transcription has made it the focus of biochemical and genetic study since its discovery⁶. DNA binding by human TFIID was first demonstrated with the adenovirus major late promoter (AdMLP).⁷ DNase I footprinting studies of the AdMLP and selected human promoters revealed

TABLE 1 Statistics of the crystallographic analysis

	Resolution (Å)	Reflections measured/unique	Completeness (%) overall/outer shell	R_{sym} (%) overall/outer shell	R_{iso}	Phasing power	R_c
(a) Data set							
MAD analysis (4 Se sites)							
λ_1 (0.9873 Å)	18.0–2.4	34,011/7,697	91.0/94.7	5.9/8.0			
λ_2 (0.9795 Å)	18.0–2.4	33,699/7,643	91.2/95.0	6.7/9.1			
λ_3 (0.9792 Å)	18.0–2.4	33,786/7,623	91.1/94.7	7.6/9.9			
λ_4 (0.9718 Å)	18.0–2.4	31,410/7,578	89.9/94.1	6.4/8.7			
Overall MAD figure of merit, 0.66							
MIR analysis							
PIP (3 Pt and 1 I sites)	15.0–2.4	62,639/8,245	98.1/98.7	5.0/8.6	0.171	0.93	0.63
Hg (2 Hg sites)	15.0–2.4	37,298/8,111	96.5/99.9	5.2/6.9	0.178	0.49	0.78
K_2PtCl_6 (2 Pt sites)	15.0–2.4	29,231/7,808	93.3/85.9	3.9/6.8	0.127	0.55	0.74
K_2PtCl_4 (2 Pt sites)	15.0–2.4	40,498/7,838	93.7/93.8	6.0/10.1	0.191	0.65	0.72
Overall MIR figure of merit, 0.51							
Overall MAD and MIR combined figure of merit, 0.73							
Native data	18.0–1.8	156,415/19,413	99.8/99.6	5.3/17.3			
(b) Refinement statistics							
	Resolution (Å)		Completeness (%)	R -factor (overall/outer shell)		Free R -factor	
Data with $ F > 2\sigma(F)$	6.0–2.0		89.8	0.198/0.218		0.244	
R.m.s. deviation	Bond lengths, 0.009 Å		Bond angles, 1.2°	Thermal parameters, 1.7 Å ²			

Crystals of the dTAF_{II}42(11–95)/dTAF_{II}62(1–82) complex containing a single mutation (Cys 55 → Ser in dTAF_{II}42) were obtained using hanging-drop vapour diffusion. The dTAF_{II}42(11–95) and dTAF_{II}62(1–82) constructs were also expressed in selenomethionine (Se–Met) substituted forms⁵⁰, and the Se–Met complex was purified as described above. A mercury adduct of the dTAF_{II}42(11–95)/dTAF_{II}62(1–82) complex was prepared by dialysing purified wild-type dTAF_{II}42(11–95)/dTAF_{II}62(1–82) complex against HgCl₂. Oscillation photography at 100K and data reduction were performed as previously described⁸. Native data were collected using Beamline X12B, at the National Synchrotron Light Source, Brookhaven National Laboratory (BNL). Se–Met MAD data and MIR data from the mercury derivative were collected at BNL using Beamline X4A. The remaining heavy-atom derivative data were collected as previously described¹¹. MAD⁵⁰ and MIR (MLPHARE; Z. Otwinowski) phase sets were calculated separately giving figure of merit (f.o.m.) values of 0.66 and 0.51, respectively. Phase combination yielded a final overall f.o.m. of 0.73 at 2.4 Å resolution, giving a high-quality experimental electron-density map revealing continuous electron density for almost the entire binary complex. Model building interspersed with positional refinement allowed an unambiguous trace and sequence assignment dTAF_{II}42(17–86) and dTAF_{II}62(1–70). The remaining 6 N-terminal and 9 C-terminal residues of dTAF_{II}42, and 12 C-terminal residues of dTAF_{II}62, were disordered and were omitted from the current model. The average B -factor for all protein atoms is 18 Å², the 125 water molecules included in the model have an average B -factor of 29 Å², and the 7 Zn²⁺ ions have an average B -factor of 28 Å². Atomic coordinates and structure factor amplitudes will be submitted to the Brookhaven Protein Data Bank. $R_{\text{sym}} = \sum |I - \langle I \rangle| / \sum I$, where I is the observed intensity, and $\langle I \rangle$ is the average intensity obtained from multiple observations of symmetry related reflections. Mean fractional isomorphous difference = $\sum ||F_{\text{PH}}| - |F_{\text{P}}|| / \sum |F_{\text{P}}|$, where $|F_{\text{P}}|$ is the protein structure factor amplitude, and $|F_{\text{PH}}|$ is the heavy-atom derivative structure factor amplitude. Phasing power = $\text{r.m.s.}(|F_{\text{H}}|/E)$, where $|F_{\text{H}}|$ is the heavy-atom structure factor amplitude and E is the residual lack of closure. R.m.s., bond lengths and r.m.s. bond angles are the respective root-mean-square deviations from ideal values. R.m.s. thermal parameter is the r.m.s.d. between the B -values of covalently bonded atomic pairs. Free R -factor was calculated with 10% of data omitted from the structure refinement. PIP, di-*m*-idobis(ethylenediamine)diplatinum(II) nitrate (Strem Chemicals).

sequence-specific interactions between TFIID and the TATA element that are primarily mediated by TBP (reviewed in ref. 8). In contrast, protection both upstream and downstream of the TATA box is largely sequence independent, displays a nucleosome-like pattern of DNase I hypersensitivity, differs between promoters (for example, AdMLP and the human heat-shock protein 70 (Hsp70) promoter protect residues –47 to +35 and –35 to –19, respectively), and is induced by some activators (reviewed in ref. 9). TATA box binding by TFIID or TBP precludes packaging of the core promoter with histone proteins (H2A, H2B, H3 and H4). Conversely, core promoter packaging by histone octamers prevents TFIID or TBP binding to the TATA element, effectively repressing transcription (reviewed in ref. 10).

Publication of the sequence of yeast TBP was followed rapidly by the sequences of homologous genes from various eukaryotes and an archaeobacterium (amino-acid identities within the carboxy-terminal portion range between 38 and 100%; reviewed in ref. 11), and considerable efforts have been directed towards understanding the mechanisms by which TBP acts in Pol II transcription. Recombinant TBP alone can bind both general and regulatory factors and direct PIC assembly and basal transcription (reviewed in refs 8, 11). Activator-dependent transcription, however, requires TBP and the remaining subunits of TFIID, the TBP-associated factors (TAF_{II}s), and some non-TAF_{II} coactivators (reviewed in ref. 9). Gene disruption studies of four yeast TAF_{II}s demonstrated that they are essential^{12,13}.

Affinity purification of TFIID allowed the identification of a set

of phylogenetically conserved TAF_{II}s, denoted by their origins and relative molecular mass (reviewed in ref. 9). In both human and *Drosophila*, TAF_{II}s are tightly associated with TBP, providing binding sites for many different transcriptional activators and coactivators that modulate transcription initiation by Pol II by means of specific protein–protein interactions with TFIID (reviewed in ref. 9). Reconstitution of TFIID has been achieved *in vitro*¹⁴, providing important insights into the roles of individual subunits as activator-specific targets that facilitate TFIID recruitment¹⁵.

Primary structure analyses of some TAF_{II}s have indicated considerable sequence homology with non-linker histone proteins (Fig. 1)^{16–19}. In *Drosophila*, dTAF_{II}42 (refs 16, 20) and dTAF_{II}62 (refs 16, 21) resemble H3 and H4, respectively, and correspond to human hTAF_{II}31 (ref. 17) and hTAF_{II}80 (refs 17, 21). Both *Drosophila* and human TFIID also contain TAF_{II}s (dTAF_{II}30 α and hTAF_{II}20)^{16,18,22,23} that are putative histone H2B homologues¹⁸, but appear to lack histone H2A homologues. The TAF_{II} nomenclatures of Roeder and Tjian (reviewed in ref. 9) have been adopted for human and *Drosophila* TFIID, respectively, the only exceptions being dTAF_{II}62 and dTAF_{II}42 which are denoted dTAF_{II}60 and dTAF_{II}40 by Tjian. We have previously documented a structural connection between eukaryotic transcription and DNA packaging. The co-crystal structure of the DNA-binding domain of the liver-specific transcription factor hepatocyte nuclear factor (HNF) 3- γ (ref. 24) resembles the structure of the linker histone H5 obtained without DNA²⁵.

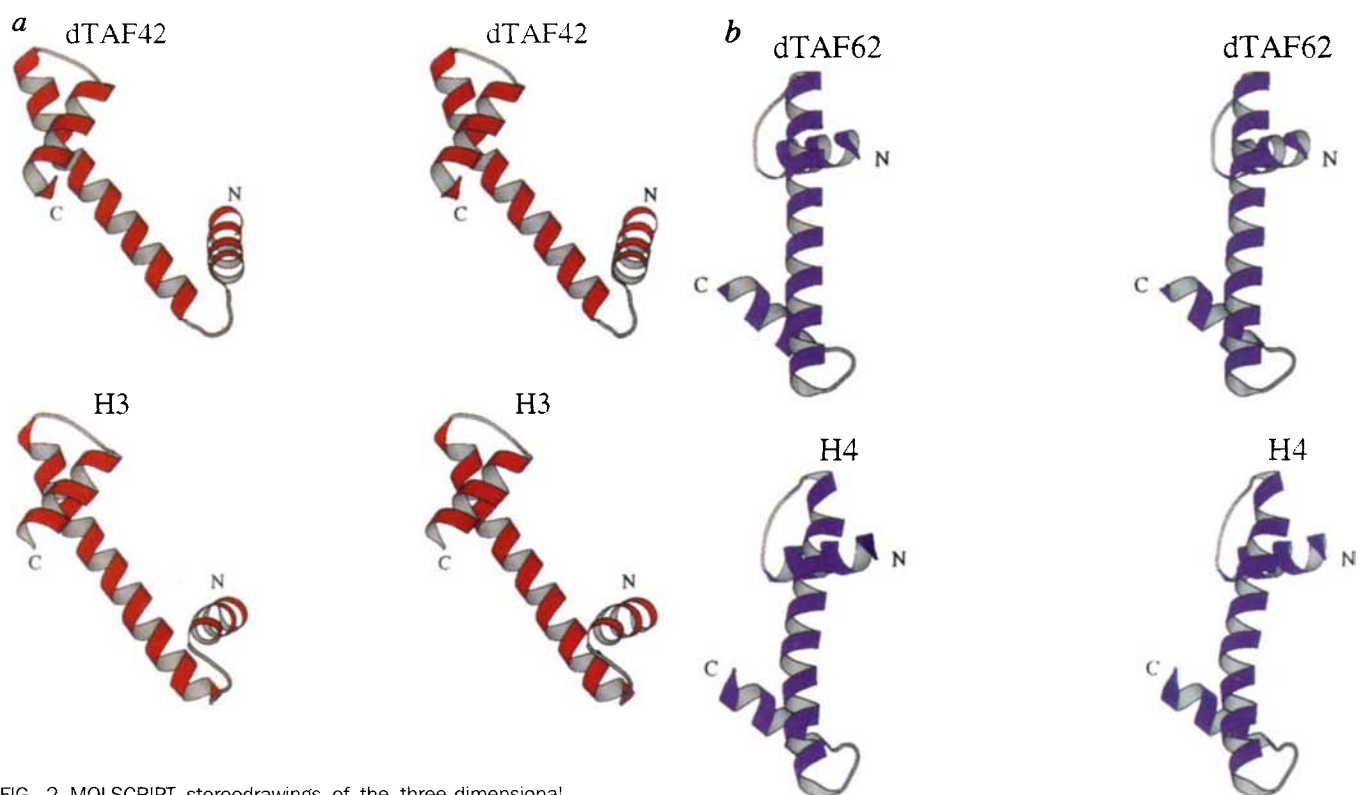


FIG. 2 MOLScripT stereodrawings of the three-dimensional structures of dTAF_{II}42(17–86) and dTAF_{II}62(1–70), and their binary complex dTAF_{II}42(17–86)/dTAF_{II}62(1–70). The corresponding views of histones H3 and H4, provided by G. Arents and E. N. Moudrianakis, have been included for comparison. Energetically favourable interactions include salt bridges (<4 Å), hydrogen bonds (<3.4 Å), van der Waals contacts (<4 Å) and two water-mediated bridges. **a**, dTAF_{II}42(17–86) (top) and H3 (bottom). The additional N-terminal helix of H3 visualized in the structure of the histone octamer core has been omitted for clarity. **b**, dTAF_{II}62(1–70) (top) and H4 (bottom). **c**, dTAF_{II}42(17–86)/dTAF_{II}62(1–70) (top) and H3/H4 (bottom). There is one water molecule trapped between the two long α -helices (see Fig. 3).

Moreover, HNF3- α , a related factor, stabilizes a precisely positioned nucleosomal array in the enhancer of the mouse albumin gene, where it may function as a sequence-specific linker histone²⁶.

Here we report the crystal structure of a complex of two RNA polymerase II-specific, TAFs from *Drosophila*. The 2.0-Å resolution structure reveals that the N-terminal portions of dTAF_{II}42 and dTAF_{II}62 are folded into the canonical histone motif. Like histones H3 and H4, dTAF_{II}42 and dTAF_{II}62 form an intimate heterodimer. Within the crystal lattice two-fold symmetry generates a dTAF_{II}42/dTAF_{II}62 dimer of heterodimers, which exists in solution and resembles the histone (H3/H4)₂ heterotetramer. We have provided what we believe to be the first high-resolution view of the histone fold, and demonstrated another structural connection between the macromolecular machines responsible for transcription and DNA packaging.

Crystallization and structure determination

Following on from the first delineation of histone-like portions of dTAF_{II}42 and dTAF_{II}62 from primary sequence alignments¹⁶, we expressed dTAF_{II}42(1–100) and dTAF_{II}62(1–91) separately as inclusion bodies in *Escherichia coli*, with the latter as a glutathione S-transferase (GST) fusion protein. Equimolar amounts of denatured dTAF_{II}42(1–100) and dTAF_{II}62(1–91) were co-renatured

and purified to homogeneity. The resulting dTAF_{II}42(1–100)/dTAF_{II}62(1–91) complex preparation was monodisperse and tetrameric, as judged by dynamic light scattering (data not shown), and largely α -helical, as judged by circular dichroism spectroscopy (data not shown).

Initial crystallization trials immediately yielded small crystals, which diffracted weakly to 4–5 Å and resisted attempts to improve

their properties. This result underscores the fact that, although monodisperse preparations of globular proteins and protein-DNA complex are excellent candidates for crystallization trials^{27,28}, monodispersity does not guarantee diffraction-quality crystals. Reasoning that flexible portions of one or both of these truncated proteins might be interfering with formation of a well-ordered crystal lattice, we searched for a proteolytic limit digest of the dTAF_{II}42(1–100)/dTAF_{II}62(1–91) complex. (For the results of a retrospective analysis of our crystallization trials with the Max-DNA complex, see ref. 29). Proteolysis was combined with mass spectrometry to rapidly obtain accurate cleavage maps²⁹ (Fig. 1a). These data suggested that three of the four termini of dTAF_{II}42(1–100) and dTAF_{II}62(1–91) had been chosen incorrectly. New *E. coli* expression vectors encoding dTAF_{II}42(11–95) and GST-dTAF_{II}62(1–82) were assembled, allowing the production of another truncated form of the dTAF_{II}42/dTAF_{II}62 complex. Crystallization trials immediately yielded large crystals in the tetragonal space group *P*4₁2₁2 (*a* = 59.8 Å, *c* = 111.0 Å), with one heterodimer in the asymmetric unit that diffracted to at least 1.4 Å resolution (Fig. 1). Trypsin treatment of (H3/H4)₂ yielded a stable complex with improved crystallization properties³⁰.

The dTAF_{II}42/dTAF_{II}62 X-ray structure was determined at 2.4 Å resolution by a combination of multiwavelength anomalous dispersion (MAD) and multiple isomorphous replacement (MIR) (Fig. 1, Table 1), followed by X-PLOR refinement to 2.0-Å resolution³¹ (Table 1). Our model consists of residues 17–86 of dTAF_{II}42, residues 1–70 of dTAF_{II}62, 125 water molecules, and 7 Zn²⁺ ions (Fig. 2), giving an *R*-factor of 19.8% with *R*_{free} = 24.4%. The electron density for the polypeptide backbone is everywhere continuous at 1.2σ in a (2|*F*_o| – |*F*_c||) difference Fourier synthesis (Fig. 3). PROCHECK³² revealed 1/140 unfavourable (*φ*, *ψ*) combinations, and main-chain and side-chain structural parameters consistently better than those expected at 2.0-Å resolu-

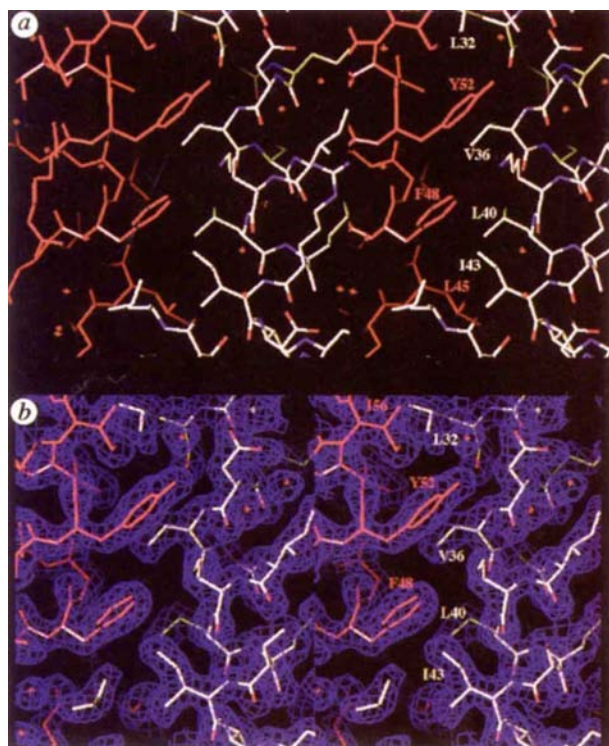


FIG. 3 a, Stereo drawing of the heterodimer interface, showing the interactions between dTAF_{II}42(17–86) and dTAF_{II}62(1–70). Critical hydrophobic residues at the interface are labelled with single-letter amino-acid code. The water molecule bridging Tyr 52 to Asp 35 and Thr 81 to Asp 35 can be seen above the OH group of Tyr 52. b, Stereo drawing of the 2.0-Å resolution (2|*F*_o| – |*F*_c||) difference Fourier synthesis for the portion of the TAF_{II} complex depicted in a.

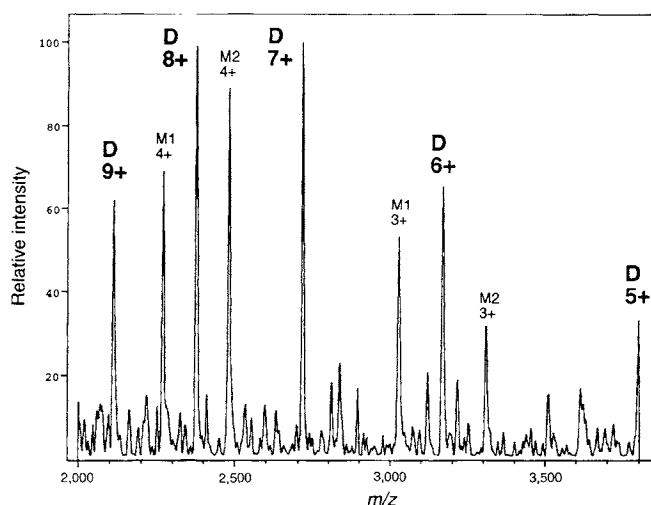


FIG. 4 Electrospray-ionization mass spectrum of the co-renatured dTAF_{II}42(11–95)/dTAF_{II}62(1–82) complex dissolved in water, showing peaks corresponding to multiply charged states of the non-covalent dTAF_{II}42(11–95)/dTAF_{II}62(1–82) heterodimer (D) plus the dTAF_{II} monomers (M1 and M2).

tion (overall *G*-factor = 0.5). A full account of the refinement at the diffraction limit will be published elsewhere.

Structures of dTAF_{II}42 and dTAF_{II}62

The X-ray structures of the histone-like regions of dTAF_{II}42 and dTAF_{II}62 are illustrated in Fig. 2. (For clarity, the TAF_{II} fragments used in this study are denoted dTAF_{II}42(11–95) and dTAF_{II}62(1–82). The portions of these two polypeptide chains visualized in the final electron-density map are denoted dTAF_{II}42(17–86) and dTAF_{II}62(1–70), with their histone-like regions denoted dTAF_{II}42(22–83) and dTAF_{II}62(9–70), respectively.) As predicted^{16,19}, the N-terminal portions of both dTAF_{II}42 and dTAF_{II}62 adopt the canonical histone motif³³, consisting of a long central α-helix flanked on each side by a short α-helix (Fig. 2). The truncated form of dTAF_{II}42 used for crystallization does not include the portion of the polypeptide chain corresponding to the additional H3 N-terminal α-helix, observed in the histone octamer structure³⁴. Throughout this paper, we will describe structural details using α1, L1, α2, L2 and α3 to denote the segments within the histone fold (italics/sloping Greek denote dTAF_{II}62). Comparisons of the histone-homology regions of the TAF_{II} structures with histones H3 and H4 revealed the following root-mean-square deviations (r.m.s.ds) between α-carbon atomic positions: dTAF_{II}42(22–83) versus dTAF_{II}62(9–70), 1.6 Å; dTAF_{II}42(22–83) versus H3(68–130), 1.6 Å; and dTAF_{II}62(9–70) versus H4(31–92), 1.6 Å, respectively. These values compare favourably with those obtained by comparing H2A, H2B, H3 and H4 with one another (see Table 1 in ref. 33). Presumably these relatively large r.m.s.ds reflect differences in the trajectory of α2. In H3 the α2 helix is linear, but is bent near its C terminus in dTAF_{II}42; the converse is true for H4 and dTAF_{II}62.

The canonical histone fold has been postulated to arise from gene duplication³³, giving two helix-strand-helix (HSH) segments (HSH1 and HSH2 refer to the N- and C-terminal halves of the histone fold, respectively). The results of a detailed comparison of HSH1 from dTAF_{II}42(22–51), dTAF_{II}62(9–38), H3(68–98) and H4(31–60), and HSH2 from dTAF_{II}42(52–83), dTAF_{II}62(39–70), H3(99–130) and H4(61–92), are shown in Fig. 1. The HSH2 fold is very similar, displaying r.m.s.ds between α-carbon atomic positions of 0.7–1.0 Å. In contrast, the HSH1 fold is less similar, with the corresponding r.m.s.ds ranging from 1.4 to 1.9 Å. This marked difference between HSH1 and HSH2 may reflect the critical role of HSH2 in mediating interactions

between heterodimers within the histone octamer crystal structure³⁴ or between TAF_{II} heterodimers and homodimers.

Differential scanning calorimetry demonstrated a single cooperative unfolding transition for the H3/H4 complex³⁵. By analogy, the histone-like portions of dTAF_{II}42 and dTAF_{II}62 are probably unable to assume their full native folds in the absence of one another, because they too demonstrate only a small number of long-range intramolecular contacts. Residues involved in these contacts have been referred to as 'self' residues³³ (denoted with 's' in Fig. 1a). For dTAF_{II}42(17–86) these intramolecular interactions include $\alpha 1$ – $\alpha 2$, $\alpha 1$ –L1, $\alpha 2$ –L1, $\alpha 2$ –L2, $\alpha 2$ – $\alpha 3$ and L2– $\alpha 3$; for dTAF_{II}62(1–70), they include $\alpha 1$ – $\alpha 2$, $\alpha 1$ –L1, $\alpha 2$ –L1, $\alpha 2$ –L2, $\alpha 2$ – $\alpha 3$ and L2– $\alpha 3$ (italics/sloping Greek are used throughout this paper to refer to dTAF_{II}62).

Structure of dTAF_{II}42/dTAF_{II}62

The structure of the dTAF_{II}42(17–86)/dTAF_{II}62(1–70) heterodimer is shown in Fig. 2. As in the H3/H4 heterodimer, also illustrated in Fig. 2, the two histone folds interact with one another in a head-to-tail fashion³³. Stabilizing contacts between dTAF_{II}42(17–86) and dTAF_{II}62(1–70) are largely hydrophobic, span the entire length of both molecules, and are conserved with H3 and H4 (Fig. 1). Binary complex formation buries about 3,390 Å² of solvent-accessible surface area (56% of the buried surface is hydrophobic, with the remainder either polar or charged). The residues involved in heterodimer contacts have been referred to as 'pair' residues³³ (denoted with 'p' in Fig. 1a). dTAF_{II}42(17–86)/dTAF_{II}62(1–70) heterodimer interactions involve each segment of the two polypeptide chains, including $\alpha 1$ – $\alpha 1$, $\alpha 1$ – $\alpha 2$, $\alpha 1$ –L2, L1– $\alpha 2$, L1–L2, L1– $\alpha 3$, $\alpha 2$ – $\alpha 3$, $\alpha 2$ – $\alpha 2$, $\alpha 1$ –L1, $\alpha 2$ – $\alpha 1$, L2–L1, L2– $\alpha 2$ and $\alpha 3$ – $\alpha 2$. Asymmetry in these interactions probably explains why H3, H4, dTAF_{II}42 and dTAF_{II}62 do not show any significant propensity to form homodimers. Electrospray ionization mass spectrometry was also used to examine the stability of the dTAF_{II}42(11–95)/dTAF_{II}62(1–82) heterodimer. Under moderately gentle electrospray conditions, a pure water solution of co-renatured dTAF_{II}42(11–95)/dTAF_{II}62(1–82) exhibits strong multiply charged peaks corresponding to the non-covalent heterodimer and the individual monomers (Fig. 4).

Structure of (dTAF_{II}42/dTAF_{II}62)₂

The structure of the (dTAF_{II}42(17–86)/dTAF_{II}62(1–70))₂ heterotetramer is shown in Fig. 5. As in the histone octamer³⁴, the symmetry axis within the TAF_{II} tetramer coincides with a crystallographic two-fold. Interactions involving the HSH2 portion of the H3 homologue, dTAF_{II}42(17–86), stabilize the heterotetramer, burying about 670 Å² of solvent-accessible surface area (48% of the buried surface is hydrophobic, with the remainder either polar or charged). These values are typical for protein–protein complexes (reviewed in ref. 36), and are consistent with the equilibrium dissociation constant of 10^{–6} M (Fig. 1). Residues involved in dTAF_{II}42(17–86)–dTAF_{II}42(17–86) interactions in the crystalline state are denoted with an asterisk in Fig. 1a, and correspond to conserved residues in H3.

We presume that the same residues (Asp 59, Val 62, Tyr 63, His 66, Leu 79, Glu 82 and Val 83) stabilize the solution form of the (dTAF_{II}42/dTAF_{II}62)₂ heterotetramer, but cannot be certain without experimental confirmation. The proteolysis–mass spectrometry results provide indirect support for this prediction. Enzymatic cleavage at three sites (Tyr 63, Arg 68 and Asp 76)

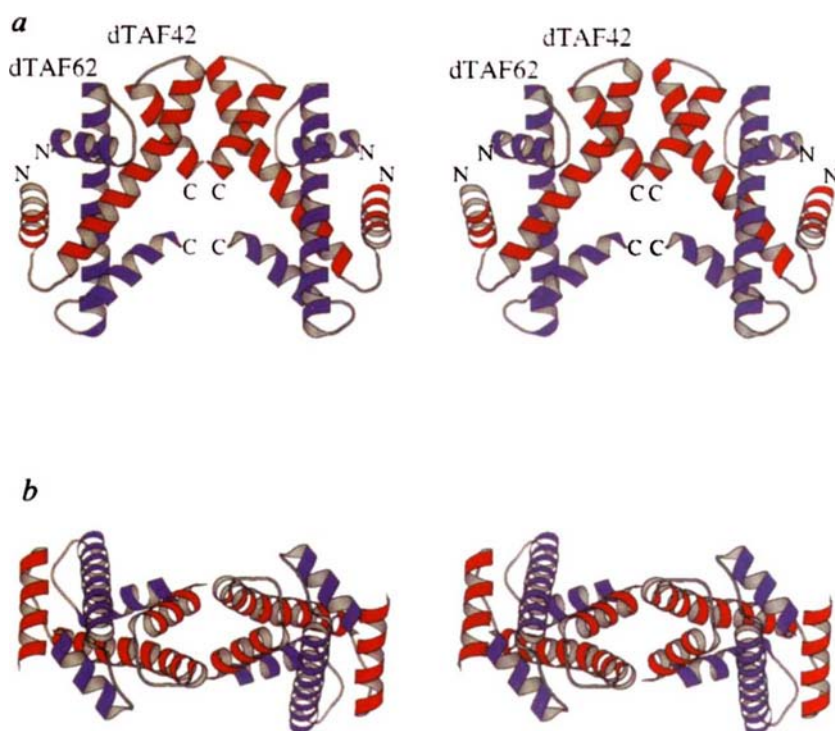


FIG. 5 Stereo drawing ribbon representations of the (dTAF_{II}42(17–86)/dTAF_{II}62(1–70))₂ heterotetramer, generated by two-fold crystallographic symmetry. *a*, View perpendicular to the two-fold symmetry axis. *b*, View along the two-fold symmetry axis.

within the HSH2 region of dTAF_{II}42(1–100) is slow compared with cleavage events outside the histone-homology region (Fig. 1a). Two of these three sites are solvent inaccessible in the crystallographic heterotetramer, and all three are completely exposed in the crystallographic heterodimer.

TFIID assembly

Our biophysical study of the dTAF_{II}42/dTAF_{II}62 complex suggests that TFIID contains a (dTAF_{II}42/dTAF_{II}62)₂ heterotetramer, which could interact with H2B-like TAF_{II}s to form a histone-like octamer. Compelling, albeit indirect, support for these predictions comes from the results of recent studies of the human TAF_{II} homologue of histone H2B (hTAF_{II}20)¹⁸. The histone-like region of hTAF_{II}20 alone allows assembly of TFIID *in vivo*, the measured hTAF_{II}20:TBP ratio in TFIID is 4 : 1, and an octamer-like pattern of protein–protein interactions has been documented for hTAF_{II}31 (homologous to H3 and dTAF_{II}42), hTAF_{II}80 (homologous to H4 and dTAF_{II}62), and hTAF_{II}20 (homologous to H2B and dTAF_{II}30 α)¹⁸. In addition, the histone-like regions of dTAF_{II}62 and dTAF_{II}30 α contain HSH2 residues homologous, respectively, to the H4 and H2B residues responsible for mediating H4–H2B interactions in the histone octamer structure³⁴ (Fig. 1a). Protein–protein interactions have also been demonstrated between hTAF_{II}80 and TBP, and between hTAF_{II}20 and TBP, hTAF_{II}135 and hTAF_{II}55 (ref. 18). Thus TFIID may contain a histone-like octameric TAF_{II} substructure ((hTAF_{II}20)₂–(hTAF_{II}31/dTAF_{II}80)₂–(hTAF_{II}20)₂ or (dTAF_{II}30 α)₂–(dTAF_{II}42/dTAF_{II}62)₂–(dTAF_{II}30 α)₂), anchored to TBP, hTAF_{II}135 (or dTAF_{II}110) and hTAF_{II}55 (there is no known *Drosophila* homologue for this human TAF_{II}).¹⁸ There are other TBP–TAF_{II} and TAF_{II}–TAF_{II} interactions (reviewed in ref. 9) that probably stabilize the remainder of the TFIID complex.

The co-crystal structure of core TFIIB recognizing the pre-formed TBP–DNA complex showed the second step of PIC assembly, explaining stabilization of the TBP–TATA element

complex by TFIIB³⁷. With TFIID instead of TBP, however, the situation may be more complicated. Affinity chromatography demonstrated that dTAF_{II}42 and its homologue hTAF_{II}31 interact with TFIIB^{20,38}. Similar methods identified interactions between hTAF_{II}80 (the human homologue of dTAF_{II}62) and two other general transcription factors TFIIF and TFIIE¹⁷. Although the functional significance of these interactions has not been established, it seems likely that they increase PIC stability, thereby facilitating recruitment of the corresponding general transcription factors. They may even do so by means of conformational changes induced in the TFIID complex by interactions with transcriptional activators or non-TAF_{II} coactivators.

Our current idea of activator-TFIID interactions suggests that the TAF_{II}s constitute a large multiprotein complex that sits on top of TBP, and integrates signals from many activators and non-TAF_{II} coactivators. Histone-like TAF_{II}-activator contacts include: interactions between the isoleucine-rich activation domain of NTF-1 and dTAF_{II}62 (ref. 14); interactions of the p53 and NFκB/p65 activation domains with dTAF_{II}42 (or its homologue hTAF_{II}31) and dTAF_{II}62 (or its homologue hTAF_{II}80)^{39,40} (M. Guermah & R. G. Roeder, unpublished observations); and interactions of the VP16 activation domain with dTAF_{II}42 (or its homologue hTAF_{II}31)^{20,38}. We do not know if any of the interactions between TAF_{II} histone homologues and various transcriptional activators map to the histone-like regions. The activators may target portions of the polypeptide chains of dTAF_{II}42 and dTAF_{II}62 that were removed to facilitate crystallization.

Another significant aspect of activator-TAF_{II} interactions concerns induced conformational changes in TFIID and TFIID-promoter complexes. Activator-induced changes have been demonstrated in TFIID-promoter complex^{41,42} manifested by downstream extension of the TFIID footprint well beyond the transcription start site that was correlated with increased recruit-

ment of other general factors. Later studies with highly purified TFIID have confirmed qualitative and quantitative effects of activators on TFIID binding, sometimes requiring TFIIA^{43,44}. Photoaffinity DNA-protein crosslinking studies have also revealed large changes in TAF_{II}-DNA contacts in response to interactions of TFIIA with promoter-bound TFIID (T. Oelgeschläger & R. G. Roeder, unpublished observations). Together, these data suggest that binding of transcriptional activators or coactivators can cause substantial rearrangements in the relative disposition of TFIID subunits and DNA. In this context, it might be imagined that activator-induced changes in TAF_{II}-DNA interactions allow the presumptive histone octamer-like substructure within TFIID to engage DNA. We speculate that such a structure might be involved in stabilizing an activator-TFIID-promoter complex, yielding a stereospecific nucleoprotein assembly that can support transcriptional activation (reviewed in ref. 45). We do not, however, predict that this putative octamer substructure would necessarily be encircled by a piece of DNA 146 base pairs long, as in the nucleosome core particle (reviewed in refs 46–49). It is also possible that TFIID merely exploits the histone fold to mediate protein-protein contacts between various TAF_{II}s.

Conclusion

We believe that the 2.0-Å resolution structure of the dTAF_{II}42/dTAF_{II}62 complex provides the first structural knowledge of RNA polymerase II TAFs, and the first high-resolution structure of the histone fold. Our work also provides a starting point for further crystallographic, biochemical and genetic studies of TFIID assembly and its role in Pol II-mediated transcription activation. Finally, studies of the three known histone-like TAF_{II}s suggest an evolutionary relationship between the histone octamer and a TFIID substructure that may be functionally significant. □

Received 13 December 1995; accepted 9 February 1996.

- Roeder, R. G. *Trends biochem. Sci.* **16**, 402–408 (1991).
- Hori, R. & Carey, M. *Curr. Opin. Genet. Dev.* **4**, 236–244 (1994).
- Zawel, L. & Reinberg, D. *Prog. Nucleic Acid Res. molec. Biol.* **44**, 67–108 (1993).
- Zawel, L., Kumar, K. & Reinberg, D. *Genes Dev.* **9**, 1479–1490 (1995).
- Koleske, A. & Young, R. *Trends biochem. Sci.* **20**, 113–116 (1995).
- Matsui, T., Segall, J., Weil, P. & Roeder, R. G. *J. biol. Chem.* **255**, 11992–11996 (1980).
- Sawadogo, M. & Roeder, R. G. *Cell* **43**, 165–175 (1985).
- Kim, J. L. & Burley, S. K. *Nature Struct. Biol.* **1**, 638–653 (1994).
- Burley, S. K. & Roeder, R. G. *A. Rev. Biochem.* **65** (in the press).
- Owen-Hughes, T. & Workman, J. L. *Crit. Rev. Euk. Gene Express.* **4**, 403–441 (1994).
- Nikolov, D. B. & Burley, S. K. *Nature Struct. Biol.* **1**, 621–637 (1994).
- Reese, J., Apone, L., Walker, S., Griffin, L. & Green, M. *Nature* **371**, 523–527 (1994).
- Poon, D. et al. *Proc. natn. Acad. Sci. U.S.A.* **92**, 8224–8228 (1995).
- Chen, J., Attardi, L., Vernijzer, D., Yokomori, K. & Tjian, R. *Cell* **79**, 93–105 (1994).
- Sauer, F., Hansen, S. & Tjian, R. *Science* **270**, 1783–1788 (1995).
- Kokubo, T. et al. *Nature* **367**, 484–487 (1994).
- Hisatake, K. et al. *Proc. natn. Acad. Sci. U.S.A.* **92**, 8195–8199 (1995).
- Hoffmann, A. et al. *Nature* **380**, 356–359 (1996).
- Baxevis, A., Arents, G., Moudrianakis, E. & Landsman, D. *Nucleic Acids Res.* **23**, 2685–2691 (1995).
- Goodrich, J., Hoey, T., Thut, C., Admon, A. & Tjian, R. *Cell* **75**, 519–530 (1993).
- Weinzierl, R., Ruppert, S., Dynlacht, B., Tanese, N. & Tjian, R. *EMBO J.* **12**, 5303–5309 (1993).
- Mengus, G. et al. *EMBO J.* **14**, 1520–1531 (1995).
- Yokomori, K., Chen, J.-L., Admon, A., Zhou, S. & Tjian, R. *Genes Dev.* **7**, 2587–2597 (1993).
- Clark, K. L., Halay, E. D., Lai, E. & Burley, S. K. *Nature* **364**, 412–420 (1993).
- Ramakrishnan, V., Finch, J., Graziano, V. & Sweet, R. *Nature* **362**, 219–223 (1993).
- McPherson, C., Shin, E.-Y., Friedman, D. & Zaret, K. *Cell* **75**, 387–398 (1993).
- D'Arcy, A. *Acta crystallogr.* **D50**, 469–411 (1994).
- Ferré-D'Amaré, A. R. & Burley, S. K. *Structure* **2**, 357–359, 567 (1994).
- Cohen, S. L., Ferré-D'Amaré, A. R., Burley, S. K. & Chait, B. T. *Protein Sci.* **4**, 1088–1099 (1995).

- Lattman, E., Burlingame, R., Hatch, C. & Moudrianakis, E. N. *Science* **216**, 1016–1018 (1982).
- Brünger, A. T. *X-PLOR, Version 3.1 manual* (Yale University, New Haven, CT, 1992).
- Laskowski, J. J., MacArthur, M. W., Moss, D. S. & Thornton, J. M. *J. appl. Crystallogr.* **26**, 283–290 (1993).
- Arents, G. & Moudrianakis, E. *Proc. natn. Acad. Sci. U.S.A.* **92**, 11170–11174 (1995).
- Arents, G., Burlingame, R. W., Wang, B.-C., Love, W. E. & Moudrianakis, E. N. *Proc. natn. Acad. Sci. U.S.A.* **88**, 10148–10152 (1991).
- Karantza, V., Friere, E. & Moudrianakis, E. N. *Biochemistry* **35** (in the press).
- Janin, J. *Proteins* **21**, 30–39 (1995).
- Nikolov, D. B. et al. *Nature* **377**, 119–128 (1995).
- Klemm, R., Goodrich, J., Zhou, S. & Tjian, R. *Proc. natn. Acad. Sci. U.S.A.* **92**, 5788–5792 (1995).
- Thut, C., Chen, J. L., Klemm, R. & Tjian, R. *Science* **267**, 100–104 (1995).
- Lu, H. & Levine, A. *Proc. natn. Acad. Sci. U.S.A.* **92**, 5154–5158 (1995).
- Horikoshi, M., Carey, M., Kadidani, H. & Roeder, R. G. *Cell* **54**, 665–669 (1988).
- Horikoshi, M., Hai, T., Lin, Y.-S., Green, M. R. & Roeder, R. G. *Cell* **54**, 1033–1042 (1988).
- Lieberman, P. & Berk, A. *Genes Dev.* **8**, 995–1006 (1994).
- Chi, T., Lieberman, P., Ellwood, K. & Carey, M. *Nature* **377**, 254–257 (1995).
- Tjian, R. & Maniatis, T. *Cell* **77**, 5–8 (1994).
- Klug, A., Rhodes, D., Smith, J., Finch, J. & Thomas, J. *Nature* **287**, 509–516 (1980).
- Richmond, T., Finch, J., Rushton, B., Rhodes, D. & Klug, A. *Nature* **311**, 532–537 (1984).
- Arents, G. & Moudrianakis, E. N. *Proc. natn. Acad. Sci. U.S.A.* **90**, 10489–10493 (1993).
- Pruss, D., Hayes, J. & Wolffe, A. *BioEssays* **17**, 161–170 (1995).
- Hendrickson, W. A. *Science* **254**, 51–58 (1991).

ACKNOWLEDGEMENTS. We thank G. Arents and E. N. Moudrianakis for access to their work before publication, help with figures, and discussion; C. Ogata for help using the X4A Beamline; M. Capel for help using the X12B Beamline; S. Ramakoti for assistance with X-ray measurements; A. Ferré-D'Amaré, J. Goldberg, H. Houbaviy, J. L. Kim, J. Kuriyan, D. B. Nikolov, G. Patikoglou and G. A. Petsko for discussions; and A. Gazes and the staff of Rockefeller University Computing Services for technical support. This work was supported in part by the Howard Hughes Medical Institute (S.K.B.), the NIH (R.G.R., B.C.T.) and the Human Frontiers Scientific Program (R.G.R.).

Discovery of interstellar dust entering the Earth's atmosphere

A. D. Taylor^{*†‡}, W. J. Baggaley[§] & D. I. Steel^{*||}

^{*} Department of Physics and Mathematical Physics, University of Adelaide, SA 5005, Australia

[†] Unit for Space Sciences, University of Kent, Canterbury, UK

[§] Department of Physics and Astronomy, University of Canterbury, Christchurch, New Zealand

^{||} Anglo-Australian Observatory, Private Bag, Coonabarabran, NSW 2537, Australia

ALL known asteroids and comets are believed to have been gravitationally bound to the Sun since they formed (together with the Sun and planets) from the solar nebula. This is because no such object has been observed with a speed exceeding the solar escape velocity, although some comets have been close to this limit¹. As comets are occasionally ejected from the Solar System, interstellar comets might be expected to arrive every few centuries, having been ejected from similar systems around other stars². The flux of interstellar dust into the Solar System should be much higher, but its detection poses significant technological challenges. Recently, the Ulysses spacecraft detected a population of dust particles near Jupiter, identified as being of interstellar origin on the basis of their speeds and trajectories^{3,4}. Here we report the radar detection of interstellar particles in the Earth's atmosphere. From intra-annual variations in particle flux, we infer the existence of two discrete sources, one associated with nearby A-type stars, and the other with the Sun's motion about the Galactic Centre. The data also suggest the presence of a third source, possibly associated with local B-type stars and young stellar clusters.

We are conducting a survey of the orbits of very faint radar meteors using the AMOR radar in New Zealand⁵. So far over 350,000 individual meteoroid orbits have been determined, many more than the total of ~69,000 from the previous surveys stored at the IAU Meteor Data Center⁶. Previous radar meteor orbit systems used Fresnel diffraction echo profiles to measure meteor speeds, and such profiles are sensitive to pulse aliasing effects, especially at high speeds, as we discuss elsewhere⁷. We determine meteor speeds from time lags between spaced receiver sites and thus can measure, with reasonable accuracy, meteor speeds exceeding the parabolic limit for orbits bound to the Solar System ($\sim 73 \text{ km s}^{-1}$; 42.1 km s^{-1} to which is added the Earth's speed and the effect of the terrestrial gravitational field).

We find that ~14% of all our orbits have in-atmosphere speeds $V > 73 \text{ km s}^{-1}$, and these we term pseudohyperbolic particles (PHPs); here we use V for the measured speeds, and v for actual speeds. Previous optical and radar meteor surveys have rendered catalogues containing up to 30% PHPs⁷. The problem lies in identifying which are true interstellar particles (ISPs) and are not just bound orbits with large measurement errors^{8,9}. Most PHPs have geocentric speeds close to the parabolic limit so that a small error in the measured speed (or radiant coordinates) would result in a hyperbolic orbit being indicated, but if real ISPs exist then some should be detected with geocentric speeds well below that limit⁸.

The results presented here use a high-quality set of 1,508 meteors with measured speeds $V > 100 \text{ km s}^{-1}$ and similar echo profiles at all receiver sites. This data set, collected between February 1990 and November 1991, comprises only 0.9% of the orbits measured in that period, owing to the stringent selection

criteria applied⁵. Comparison between Fresnel and time-lag speed determinations for meteors with $V < 50 \text{ km s}^{-1}$ leads to an estimated uncertainty $\sigma \approx 6 \text{ km s}^{-1}$ in the speeds near $V = 73 \text{ km s}^{-1}$, a value substantiated by inspection of the measured speeds in the $\sim 65 \text{ km s}^{-1}$ η -Aquarid shower^{5,7}. The data set therefore comprises meteors with speeds $> 3\sigma$ above any possible bound Solar System orbit. This means that an ISP whose real (not measured) geocentric speed was $v = 90 \text{ km s}^{-1}$ would not have been gathered into the data set used here unless its speed was erroneously measured as being $V > 100 \text{ km s}^{-1}$; thus the majority of ISPs were probably excluded from the data set analysed but this was necessary in order to exclude mis-measurements of meteoroids from bound orbits. The mean speed for this ISP data set is $\bar{V} \approx 164 \text{ km s}^{-1}$ with an uncertainty of $\pm 30 \text{ km s}^{-1}$, in agreement with a crude estimate of $\bar{V} \approx 140 \text{ km s}^{-1}$ derived solely from their mean ionization height⁷.

The meteoric ionization produced is a strong function of speed, so ISPs near our limiting magnitude ($M_{\text{lim}} = 12.5$) will be smaller than the typically $\sim 100 \mu\text{m}$ meteoroids from bound orbits we detect. Verniani¹⁰ determined an empirical relation $q = 0.0509 m^{0.92} v^{3.91}$ between the mass (m in g), the speed (v in m s^{-1}), and the electron line density (q , electrons per metre). Our M_{lim} corresponds to $q = 2.5 \times 10^{11} \text{ m}^{-1}$. At $v = 100 \text{ km s}^{-1}$ the diameter of the smallest detectable meteoroid (assumed spherical with density 1 g cm^{-3}) is $\sim 40 \mu\text{m}$; at $v = 200 \text{ km s}^{-1}$ it is $\sim 15 \mu\text{m}$. Far-infrared observations require a substantial $\sim 30\text{-}\mu\text{m}$ interstellar grain component¹¹, and Pioneer 10 and 11 dust impact data, indicating a constant flux at heliocentric distances from 3 to 18 AU, have been interpreted as being due to $\sim 10\text{-}\mu\text{m}$ interstellar grains penetrating the Solar System⁴. Interstellar particles of order tens of micrometres are therefore not unexpected.

Our data set displays strong seasonal variations (Fig. 1), arguing against these meteors being merely Solar System particles for which exceptional errors accumulated. The strongest evidence for these particles having come from interstellar trajectories lies in the relative absence of detections between days 260 and 350, when the Earth is moving away from the apex of the solar motion. The trends seen in Fig. 1 may be explained as being due to seasonally changing relative geometries between the terrestrial motion vector and discrete source directions. We tentatively identify a bimodal distribution with peaks near days 32 and 170. The raw

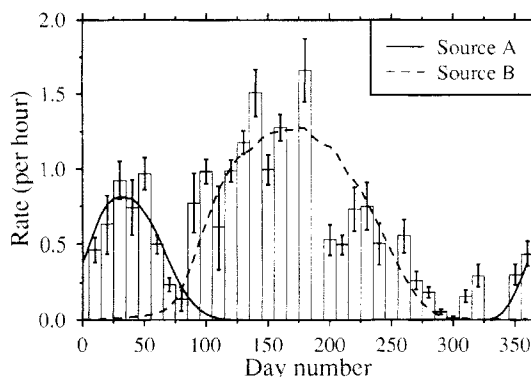


FIG. 1 Detection rates of meteors with observed velocities $V > 100 \text{ km s}^{-1}$ in ten-day bins (histogram). Uncertainty bars indicate the square roots of the count statistics. Expected distributions from two models of discrete interstellar sources are shown: ecliptic latitude $\beta = 30^\circ$, longitude $\lambda = 240^\circ$, and initial speed $v_\infty = 40 \text{ km s}^{-1}$ (source A, peak at day of year 32; solid line), and $\beta = 30^\circ$, $\lambda = 0^\circ$, $v_\infty = 80 \text{ km s}^{-1}$ (source B, peak at day 170; dashed line). As discussed in the text, an alternative possibility is that there is a source B' giving a peak near day 150, and another source C resulting in a peak near day 224.

‡ To whom correspondence should be addressed at Adelaide.

observational data render mean time lags for meteors detected between days 340 and 80 which are significantly shorter than those between days 80 and 280. The difference is equivalent to $\sim 20 \text{ km s}^{-1}$, suggesting a slower ISP source for the peak at day 32 than that at day 170.

We developed a model of the ISP influx for comparison with the observations, taking into consideration gravitational focusing by the Sun, flux enhancements resulting from the Earth's orbital motion, and increased ionization production by high-speed meteoroids. This renders an estimate of the fraction of meteors detected with $V > 100 \text{ km s}^{-1}$. We assume that ISPs flow into the Solar System from a direction defined by its ecliptic latitude (β) and longitude (λ) with a single initial speed (v_∞). Electromagnetic and radiation effects are insignificant for these particle sizes. Gravitational focusing increases the spatial density over that in interstellar space, and concentrates ISPs behind the Sun compared to positions between the Sun and the source. Figure 2a shows the spatial density at 1 AU as a function of the terrestrial ecliptic longitude (L) with respect to the influx for a hypothetical source emanating from $\beta = 30^\circ$, $\lambda = 0^\circ$, with $v_\infty = 40 \text{ km s}^{-1}$. A northern radiant was chosen because the solar motion is towards a point well north of the ecliptic (see later). Due to the geographical latitude and antenna patterns of our radar, most of our meteor radiants are in the declination range $+5^\circ > \delta > -30^\circ$ (ref. 5); thus only ISPs from low northern latitudes can appear in our data, so we chose $\beta = 30^\circ$ for our source models. Varying β in the range $0-60^\circ$ makes little difference to the results. The model particles are then propagated to the Earth, the geocentric velocity of each calculated, and a normalized flux derived. When $180^\circ <$

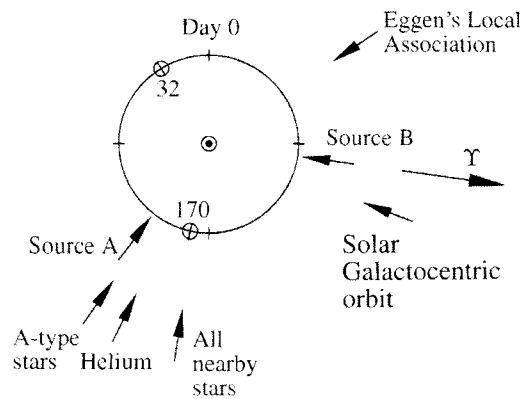


FIG. 3 Summary of possible sources of interstellar meteoroids and the influx observed, projected onto the ecliptic. The symbol γ shows the vernal equinox. Days of year 32 and 170 correspond to peaks in the count rates shown in Fig. 1 and are marked on the Earth's orbit. These peaks are explicable by our model sources A ($\sim 110^\circ$ in longitude from day 32) and B ($\sim 95^\circ$ in longitude from day 170); the differential longitude shifts are due to the different speeds used for A and B. Source A may plausibly be associated with the motion of the Solar System relative to nearby A-type stars¹³ rather than the motion relative to all other local stellar types¹². Source B is near-aligned with the Sun's Galactocentric orbit¹⁴, but temporal gaps in the data used here do not allow us to exclude the possibility that the larger peak fitted in Fig. 1 actually comprises two separate components, one (source B') centred near day 150 and then more closely associated with the solar Galactocentric orbit, whilst the other (source C) centred near day 225, is derived from Eggen's local association of star clusters and B-type stars^{15,16}.

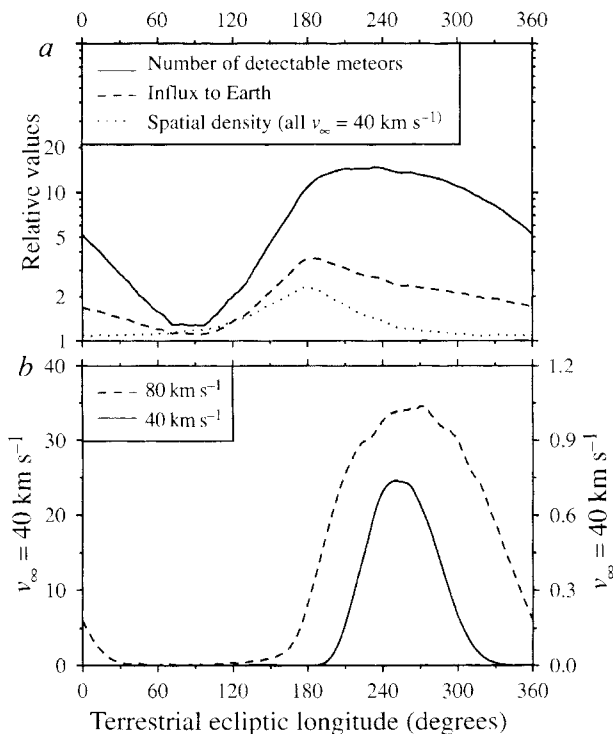


FIG. 2 a, The relative spatial density, flux to the Earth, and number of meteors above a given detection threshold, for a hypothetical interstellar source emanating from $\beta = 30^\circ$, $\lambda = 0^\circ$ with $v_\infty = 40 \text{ km s}^{-1}$. While the spatial density is enhanced by gravitational focusing behind the Sun (terrestrial ecliptic longitude $L = 180^\circ$), most meteors would be detected during the half-year as the Earth approaches the interstellar influx ($180^\circ < L < 360^\circ$). b, The relative numbers of interstellar meteors expected to be detected with measured speeds $V > 100 \text{ km s}^{-1}$ for two source models, one with $v_\infty = 40 \text{ km s}^{-1}$, as in a, and the other emanating from the same direction but with $v_\infty = 80 \text{ km s}^{-1}$.

$L < 360^\circ$ the terrestrial motion is towards the source, producing the flux enhancement seen in Fig. 2a.

For any limiting line density q and specified v , a minimum mass $m_{\min}$ producing a detectable meteor may be derived. The number of detections is $N(m > m_{\min}) = km_{\min}^{-\alpha}$; α is the cumulative mass index and k a constant. We cannot determine α from our ISP sample owing to speed measurement uncertainties and the speed-dependence of ionization production, so we adopt $\alpha = 0.6$ as representative of ISPs. Varying α has little effect on the relative shape of different velocity populations and does not affect the longitude L at which peak influxes are anticipated. The value of α does, however, affect the relative strength of sources with different v_∞ , making any estimate of that speed very model-sensitive.

The uppermost curve in Fig. 2a shows the relative numbers of ISPs that would be detected as a function of L . As the Earth approaches incoming ISPs ($L = 270^\circ$), the geocentric speeds will generally be $\sim (60 \cos \beta) \text{ km s}^{-1}$ greater than six months later while moving away from the source direction. For a heliocentric speed of 58 km s^{-1} at 1 AU (corresponding to $v_\infty = 40 \text{ km s}^{-1}$), this variation represents a factor of 3 in the geocentric speeds and thus a change in detectable mass by a factor of ~ 100 , leading to $\sim 100^\alpha \approx 16$ times more ISPs being detected. Most of these, though, would have $V < 100 \text{ km s}^{-1}$ and so would be rejected from our data set. Figure 2b shows only those test ISPs with $V > 100 \text{ km s}^{-1}$; note that the terrestrial encounter speeds are actually $v \approx 90 \text{ km s}^{-1}$, but we have applied random experimental errors in the model to deduce how many of those test particles would have $V > 100 \text{ km s}^{-1}$ measured. For $v_\infty = 40 \text{ km s}^{-1}$, the model predicts a peak detection rate when $L = 250^\circ$ ($\lambda - L = 110^\circ$). For a source in the direction but with $v_\infty = 80 \text{ km s}^{-1}$ the peak occurs at $L = 265^\circ$ ($\lambda - L = 95^\circ$) and is much broader, although there remains an absence of detections around $L = 90^\circ$. Earth encounter speeds are $v \approx 120 \text{ km s}^{-1}$ for ISPs from this source detected near the peak. If $v_\infty = 80 \text{ km s}^{-1}$ then ISPs are 47 times more likely to be detected and counted within our data set than those with $v_\infty = 40 \text{ km s}^{-1}$ ($\alpha = 0.6$); for $\alpha = 0.2$ or 1.0 the ratios are 27 and 80 times, respectively. ISPs with $v_\infty = 20 \text{ km s}^{-1}$ are over 100 times less likely to be selected

than if $v_{\infty} = 40 \text{ km s}^{-1}$. Clearly high-speed interstellar sources are strongly favoured by the speed selection ($V > 100 \text{ km s}^{-1}$) imposed here.

Next we compare these model calculations with our data set. $L \approx 100^\circ$ on 1 January, allowing the day-of-year to be easily deduced from L . In Fig. 1 we model a fit to the data with two populations, emanating from $\beta = 30^\circ$, $\lambda = 240^\circ$ with $v_{\infty} = 40 \text{ km s}^{-1}$ (source A), and from $\beta = 30^\circ$, $\lambda = 0^\circ$ with $v_{\infty} = 80 \text{ km s}^{-1}$ (source B); these are the hypothetical sources plotted in Fig. 2b differentially shifted in longitude and scaled by factors of 1.1 (source A) and 0.037 (source B) to fit the observational data. The scaling suggests that the spatial density in interstellar space of source A is $\sim 10^3$ times greater than that of Source B.

Figure 3 provides a graphical summary of these source directions, which require identification with plausible producers of ISPs. The solar motion relative to nearby stars¹² is directed towards a galactic latitude $b = 23^\circ$, longitude $l = 52^\circ$ (that is $\beta = 50^\circ$, $\lambda = 269^\circ$), near the location of source A. For late A-type stars only¹³, the apex is towards $\lambda \approx 243^\circ$, even closer to source A, and would produce a peak near day 34. Such stars are excellent candidates as ISP producers because they possess protoplanetary particulate disks, collisions and radiation pressure leading to dust ejection. Source A is also close to the flow of neutral helium through the Solar System, as associated with the $\sim 0.4\text{-}\mu\text{m}$ interstellar grains detected with Ulysses⁴, but the ISPs we have detected are too large to have come into equilibrium with that gas.

We now consider source B, associated with the faster ISPs. These ISPs might come from a direction associated with the solar galactocentric orbit¹⁴, thus emanating from $b = 0^\circ$, $l = 90^\circ$ ($\beta = 60^\circ$, $\lambda = 347^\circ$). That direction would give a peak detection rate at day 154; the observed peak is near day 170 for source B

(Figs 1 and 3). Alternatively this source might comprise two discrete components: B' centred at day 150 and C centred at day 225 (compare Fig. 1). Eggen^{15,16} describes a local association of young star clusters and B-type stars moving past the Sun. These would produce ISPs coming from $\beta = 19^\circ$, $\lambda = 121^\circ$, and a peak detection rate around day 224.

Our analysis of stringently selected radar-detected meteors demonstrates that there is a near-Earth flux of small meteoroids (sizes $\sim 15\text{--}40 \mu\text{m}$) arriving from interstellar space; these meteoroids originate from a few discrete sources in the vicinity of the Sun. We plan further data collection, and analysis of data in hand, to refine our knowledge of plausible source directions for these particles. $\square$

Received 18 July 1995; accepted 1 February 1996.

1. Kresák, Ľ. *Astr. Astrophys.* **259**, 682–691 (1992).
2. Hughes, D. W. J. *Br. astr. Assoc.* **101**, 119–120 (1990).
3. Grün, E. et al. *Nature* **362**, 428–430 (1993).
4. Grün, E. et al. *Astr. Astrophys.* **286**, 915–924 (1994).
5. Baggaley, W. J., Bennett, R. G. T., Steel, D. I. & Taylor, A. D. *Quart. J. R. astr. Soc.* **35**, 293–320 (1994).
6. Lindblad, B. A. & Steel, D. I. in *Asteroids, Comets, Meteors 1993* (eds Milani, A., Di Martino, M. & Cellino, A.) 497–501 (IAU Symp. 160, Kluwer, Dordrecht, 1994).
7. Taylor, A. D., Baggaley, W. J., Bennett, R. G. T. & Steel, D. I. *Planet. Space Sci.* **42**, 135–140 (1994).
8. Jones, J. & Sama, T. *Bull. astr. Inst. Czechoslov.* **36**, 103–115 (1985).
9. Hajduková, M. Jr in *Meteoroids and their Parent Bodies* (eds Štohl, J. & Williams, I. P.) 61–64 (Slovak Acad. Sci., Bratislava, 1993).
10. Verniani, F. J. *geophys. Res.* **78**, 8429–8462 (1973).
11. Rowan-Robinson, M. *Mon. Not. R. astr. Soc.* **258**, 787–799 (1992).
12. Jaschek, C. & Valbousquet, A. *Astr. Astrophys.* **255**, 124–129 (1992).
13. Jaschek, C. & Valbousquet, A. *Astr. Astrophys.* **242**, 77–84 (1991).
14. Mihalas, D. & Binney, J. *Galactic Astronomy: Structure and Kinematics* (Freeman, San Francisco, 1981).
15. Eggen, O. J. *Mon. Not. R. astr. Soc.* **204**, 377–390 (1983).
16. Eggen, O. J. *Mon. Not. R. astr. Soc.* **204**, 391–403 (1983).

ACKNOWLEDGEMENTS. This work was supported in part by the Australian Research Council.

Semiconducting superlattices templated by molecular assemblies

Paul V. Braun*, Paul Osenar* & Samuel I. Stupp*†‡

Departments of * Materials Science and Engineering and † Chemistry, and Beckman Institute for Advanced Science and Technology and Materials Research Laboratory, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA

ORGANIC-INORGANIC nanostructured composites provide a rich source of new materials^{1–14} for a host of technological applications. For example, the incorporation of organic molecules in an inorganic lattice can toughen an otherwise brittle material^{15–17}, or be used to tailor its electronic properties¹⁴, and cooperative interactions between organic and inorganic molecules are being used to generate a range of porous materials for separation and catalytic technologies^{4–10}. Here we describe the growth of stable semiconductor-organic superlattices based on cadmium sulphide and cadmium selenide. The template for the structures is provided by a liquid-crystalline phase formed from non-ionic organic amphiphiles, water and precursor ions for the inorganic semiconductor. Precipitation of the organic-inorganic solid takes place within the ordered environment of the mesophase, and both the symmetry and long-range order of the liquid crystal are preserved. We anticipate that materials of this type can be tailored, through the electronic properties of the organic amphiphiles, for photosynthetic and photocatalytic applications.

‡ To whom correspondence should be addressed.

A promising route to nanostructured composites uses the ordered environment of a preformed amphiphile-solvent mesophase to nucleate and grow inorganic structures. In such a system, the ordered assembly of organic molecules in the mesophase partitions the solvent and may therefore confine the reactive species, providing a controlled environment for the formation of nanostructures. In the few experiments that have attempted to harness the order of a mesophase in the synthesis of inorganic materials, precipitates with unique morphology were obtained, including oblong crystallites¹¹ and macroporous crystalline networks¹². The mesoporous oxide methodology has been applied to a concentrated system in a recent study¹³ in which the mesoporous oxide was derived from an isotropic liquid (formed by adding tetramethyl orthosilicate to a mesophase). Here we report the growth of a non-oxide, nanocomposite material templated directly by a hexagonal mesophase yielding a superlattice with matching symmetry and periodicity (Fig. 1). We believe this is the first example of a reaction leading to a nanostructure templated with the characteristic spacings and symmetry of the precursor mesophase. In our system the nature of the mesophase does not change at any time throughout the synthetic pathway leading from precursor to product. This is in contrast to earlier work based on the cooperative assembly of condensing oxide precursors with amphiphilic molecules, in which nanocomposites were precipitated from dilute solutions or isotropic solutions formed by adding precursor to a mesophase^{6–10,13}. However, in none of these cases did the precipitation take place in the preordered environment of a mesophase.

We have targeted the synthesis of nanocomposites containing II–VI semiconductors. The organized, molecular dispersion of organic in a semiconductor lattice could provide materials with properties not attainable by conventional techniques, as the intimate association of organic and inorganic components in a

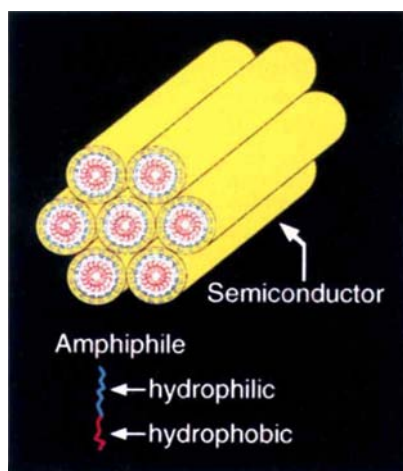


FIG. 1 Schematic representation of an ordered nanocomposite solid in which the organic phase consists of hexagonally close-packed tubules of self-assembled amphiphiles. An inorganic semiconductor (yellow regions) precipitates in the hydrophilic regions of the self-assembled structure with the diffusion of cadmium ions from the surrounding area. Precipitation results in ~ 100 -nm nanostructured particles distributed throughout the mesophase.

regular array should profoundly affect the electrical and chemical properties. To our knowledge, such structures have not previously been described, but efforts have been reported on the synthesis of nanoscale CdS. These synthetic methodologies include precipitation in a variety of media^{18–28}, but they all produce solid semiconductor particles with morphologies never very far from spherical (with varying degrees of control over the particle diameter). CdS has also been synthesized under Langmuir monolayers, sometimes with a dendritic structure^{19,29,30}.

The nanostructure of cadmium sulphide synthesized by precipitation in the hexagonal mesophase shows hexagonal symmetry on a length scale commensurate with that of the mesophase (Figs 2 and 3). The mesophase consists of cylindrical aggregates of amphiphilic molecules which pack in a hexagonal array. The cadmium ions must reside in the hydrophilic skin of the cylinders, therefore we believe that the cavities in Fig. 2 represent the space occupied by the hydrophobic core of oleyl molecular segments. In the mesophase, X-ray diffraction shows the distance between the centres of the cylinders to be ~ 8 nm, which matches the centre-to-

centre distance between cavities. In addition, molecular models show the overall extended length of the amphiphile to be 5.8 nm, with a hydrophobic segment roughly 2.2 nm in length. We expect the amphiphile to be in a less than extended conformation, with an overall length of approximately half that of the centre-to-centre distances between cavities. The conformation of the amphiphile shown in Fig. 2 would be compatible with the superlattice dimensions as measured by electron microscopy, and also with the periodicity of the original mesophase measured by X-ray diffraction. Remarkably, the symmetry and dimensions of the mesophase persist throughout the synthesis, indeed the mesophase is even preserved as by-products of the mineralization reaction are formed, as is shown clearly in the X-ray diffraction patterns in Fig. 4.

The observed nanostructured particles are not the product of flocculation of nanoscale-diameter spheres. If the observed structure consisted of aggregates of such nanoscale spheres, individual spheres would have been observed by transmission electron microscopy after repeated sonication of the product. An example of such aggregates, as well as their constituent nanospheres, was observed in ref. 31. Further evidence for the monolithic nature of our nanostructured particles is provided by the visible absorption spectrum, which corresponds to that of bulk CdS rather than the individual nanoscale particles in which quantum confinement is expected.

When CdS is synthesized from mesophases doped with cadmium chloride, the product consists of particles 5 to 10 times larger than those obtained using cadmium acetate. Indeed, the product is also a nanostructured solid with a hexagonal superlattice of cavities. The formation of large, nanostructured particles is possibly the result of coarsening, which can occur in this case by dissolution of CdS by HCl:



Ageing this product in the mesophase for one week before collecting the particles resulted in even larger-diameter nanostructured particles. This suggests that an Oswald ripening process is taking place in the system. In great contrast, CdS produced from cadmium acetate does not coarsen because the inorganic lattice is insoluble in acetic acid, and so the back reaction does not take place. In both cases, of course, the equilibrium is driven to the right by excess H_2S . The possibility that particles are indeed coarsening with superlattice morphology implies thermodynamic stability in a composite solid. If the superlattice structure were kinetically trapped, HCl would have dissolved it, and ripening would produce a different morphology on the edges of the particles.

FIG. 2 High-magnification electron micrograph of the semiconductor superlattice with a cross-section schematic overlay of the templating molecules drawn roughly to scale. Note the match in electron density contrast with respect to the overlay, showing dense regions in hydrophilic domains (due to presence of CdS) and lighter regions in the position of hydrophobic cores. The crystalline nature of the product is indicated by the lattice fringes observed towards the right of the micrograph. Centre-to-centre distances between cavities where there does not seem to be any cadmium sulphide are ~ 8 nm, and the cavities themselves have diameters of 3 nm. On the right is a representation of the amphiphile, generated through a molecular dynamics program, with a hydrophilic segment (blue) ~ 2.5 nm in length, and a hydrophobic segment (red) ~ 1.5 nm in length. This conformation is reasonable given that neither fully extended nor folded conformers are expected in the hexagonal mesophase. The overlay on the top left of the micrograph shows how the amphiphile might be anchored in the CdS lattice.

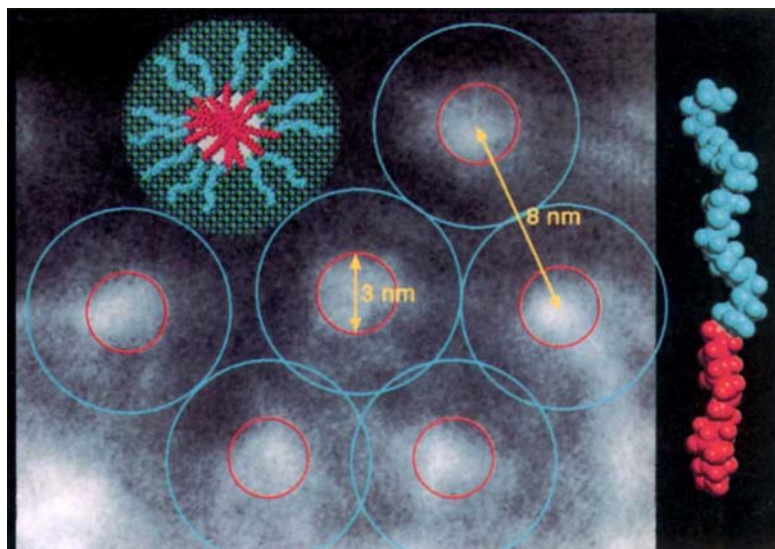
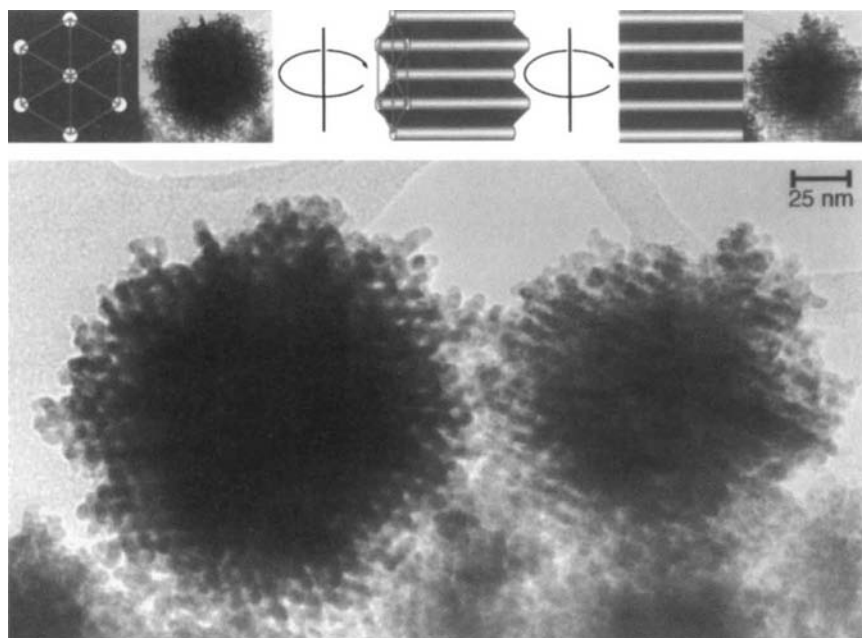


FIG. 3 The left of the transmission electron micrograph shows nanostructured CdS viewed with the cylindrical amphiphile aggregates perpendicular to the micrograph surface; the one on the right shows the cylindrical structures parallel to the surface. The inset shows a model of the nanostructure as a hexagonal arrangement of cylindrical pores of low electron density corresponding to organic material in a solid matrix of semiconductor. The left and right schematic representations correspond to the micrograph; the central view shows an intermediate rotation of the cylindrical assemblies. CdS was precipitated by flowing hydrogen sulphide gas over a hexagonal mesophase doped with a cadmium diacetate dihydrate. The mesophase was formed using oligoethylene oxide oleyl ether with an average of ten ethylene oxide units ($C_{18}H_{35}(OCH_2CH_2)_{10}OH$) mixed with an equal volume of aqueous 0.1 M cadmium salt. The diffusion of H_2S into the mesophase converted the cadmium salt to bright yellow CdS and acetic acid by-product. A sample of mesophase 10 mm thick reacts completely in ~ 8 h. The resulting CdS-mesophase composite was washed repeatedly with diethyl ether/ethanol (50/50) to remove by-products and unbound amphiphile, and then dried under high vacuum. Transmission electron micrograph samples were prepared by suspending the precipitate in ethanol, then casting on holey carbon.



Elemental analysis indicated that the organic content in the nanostructured particles produced from cadmium acetate is about 15% by volume, even after refluxing in a good solvent for the amphiphile (ethanol). As a control, CdS was also grown from cadmium acetate in water, and the resulting spherical particles, of diameter 10–15 nm, were then exposed to the amphiphile in aqueous solution. After refluxing in the same solvent, the CdS particles had an organic content of only 5% by volume, despite their larger surface-to-volume ratio relative to the nanostructured material. We conclude from these observations that amphiphilic molecules are anchored in the CdS during precipitation, and that hydrophobic molecular segments must be populating the cavities of the nanostructured material. The product of our synthesis could

therefore be described as a composite solid in which organic structures are molecularly dispersed in the inorganic lattice.

We attribute the superlattice morphology of the nanocomposite to a templating effect of the hexagonal mesophase. The templating mechanisms are not yet fully understood, but we believe that various factors are involved. One such factor is the confinement of ions to the hydrophilic region of the mesophase and their direct association with the polar moieties of the amphiphile. The importance of ion–amphiphile interactions in the formation of a CdS superlattice structure was indicated by our attempts to synthesize the analogous composites with PbS and CdSe. In the PbS case only large cubic crystals are formed, but when we synthesized the CdSe system a nanostructure commensurate with the original mesophase was again observed. Thus the specific cation–ethylene oxide interaction, both in the doped mesophase and in the precipitated semiconductor, must be a crucial factor in the formation of a templated superlattice. Two additional factors that could influence the formation of a superlattice are the reduction of surface energy in the inorganic lattice by molecularly dispersed organic molecules, and the thermodynamic stability of the templating mesophase. With regard to this last factor, mesophases of similar structure but lower isotropization temperatures failed to produce the superlattice morphology.

The morphology and composition of the nanocomposite solid discovered here suggest various technological possibilities. One interesting potential of these structures is the possibility of using the organic component, anchored in the semiconductor, as a tool to impact electronic properties. For example, a liquid crystal-forming amphiphile with electron or hole accepting capability could mediate electron-hole recombination rates and chemical reactivity. This could lead to new photosynthetic or photocatalytic materials, useful for conversion of solar energy. Furthermore, the ability to generate a periodic array of nanoscale pores in a semiconductor continuum could provide an opportunity to access the interesting electronic behaviour of antidots. Cylindrical channels lined with organic segments may also provide selective transport for molecules to be exposed to transformations within the semiconductor. □

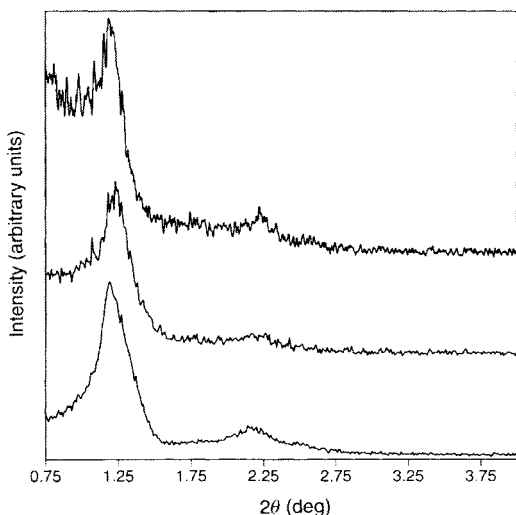


FIG. 4 X-ray diffraction revealing hexagonal order in the pure mesophase (bottom trace), the mesophase containing 0.1 M cadmium acetate (middle trace), and the H_2S -saturated mesophase containing nanostructured CdS particles and the reaction by-product (acetic acid) (top trace). Similar birefringent textures were seen at all these stages by polarized optical microscopy.

Received 7 August 1995; accepted 5 February 1996.

1. Archibald, D. D. & Mann, S. *Nature* **364**, 430–433 (1993).
2. Meldrum, F. C., Wade, V. J., Nimmo, D. L., Heywood, B. R. & Mann, S. *Nature* **349**, 684–687

- (1991).
- Meldrum, F. C., Heywood, B. R. & Mann, S. *Science* **257**, 522–523 (1992).
 - Burkett, S. L. & Davis, M. E. *Chem. Mater.* **7**, 920–928 (1995).
 - Szostak, R. *Handbook of Molecular Sieves* (Van Nostrand Reinhold, New York, 1992).
 - Kresge, C. T., Leonowicz, M. E., Roth, W. J., Vartuli, J. C. & Beck, J. S. *Nature* **359**, 710–712 (1992).
 - Beck, J. S. et al. *Chem. Mater.* **6**, 1816–1821 (1994).
 - Monnier, A. et al. *Science* **261**, 1299–1303 (1993).
 - Huo, Q. et al. *Nature* **368**, 317–321 (1994).
 - Firouzi, A. et al. *Science* **267**, 1138–1143 (1995).
 - Friebreg, S. E. & Wang, J. J. *Dispers. Sci. Technol.* **12**, 387–402 (1991).
 - Walsh, D., Hopwood, J. D. & Mann, S. *Science* **264**, 1576–1578 (1994).
 - Attard, G. S., Glyde, J. C. & Göttner, C. G. *Nature* **378**, 366–368 (1995).
 - Mitzi, D. B., Feild, C. A., Harrison, W. T. A. & Guloy, A. M. *Nature* **369**, 467–469 (1994).
 - Berman, A., Addadi, L. & Weiner, S. *Nature* **331**, 546–548 (1988).
 - Berman, A. et al. *Science* **250**, 664–667 (1990).
 - Berman, A. et al. *Science* **259**, 776–779 (1993).
 - Steigerwald, M. L. & Brus, L. E. *Acc. Chem. Res.* **23**, 183–188 (1990).
 - Fendler, J. H. *Membrane-Mimetic Approach to Advanced Materials* (Springer, Berlin, 1994).
 - Dameron, C. T. et al. *Nature* **338**, 596–597 (1989).
 - Wang, Y. & Herron, N. J. *phys. Chem.* **95**, 525–532 (1991).
 - Bianconi, P. A., Lin, J. & Strzelecki, A. R. *Nature* **349**, 315–317 (1991).
 - Lin, J., Cates, E. & Bianconi, P. A. *J. Am. chem. Soc.* **116**, 4738–4745 (1994).
 - Yuan, U., Fendler, J. H. & Cabasso, I. *Chem. Mater.* **4**, 312–318 (1992).
 - Cummins, C. C., Schrock, R. R. & Cohen, R. E. *Chem. Mater.* **4**, 27–30 (1992).
 - Moffitt, M. & Eisenburg, A. *Chem. Mater.* **7**, 1178–1185 (1995).
 - Moffitt, M., McMahon, L., Pessel, V. & Eisenburg, A. *Chem. Mater.* **7**, 1185–1192 (1995).
 - Weller, H. *Angew. Chem.* **32**, 41–53 (1993).
 - Yang, J., Meldrum, F. C. & Fendler, J. H. *J. phys. Chem.* **99**, 5500–5504 (1995).
 - Facci, P., Erokhiin, V., Tronin, A. & Nicolini, C. *J. phys. Chem.* **98**, 13323–13327 (1994).
 - Brust, M., Bethell, D., Schiffrin, D. J. & Kiely, J. K. *Adv. Mater.* **7**, 795–797 (1995).

ACKNOWLEDGEMENTS. We thank S. Kennedy for experimental assistance, and M. Kaser for help with molecular graphics. This work was supported in part by a grant from the Department of Energy Center of Excellence for the Synthesis and Processing of Advanced Materials. Partial support of this work was obtained from the Beckman Institute for Advanced Science and Technology, through a Beckman assistantship for P.V.B.

A self-consistent phase diagram for supercooled water

Hideki Tanaka

Division of Polymer Chemistry, Graduate School of Engineering, Kyoto University, Sakyo, Kyoto 606-01, Japan

MANY of the low-temperature thermodynamic properties of water, such as heat capacity and isothermal compressibility, exhibit anomalous behaviour that tends to diverge in the supercooled state^{1,2}. On the basis of molecular dynamics simulations^{3,4}, these phenomena have recently been attributed to the influence of a critical point terminating the temperature–pressure coexistence line separating low- and high-density amorphous ices. But the fact that water tends to lose its anomalous behaviour⁵ in the pressure range predicted for the new critical point poses problems for such an interpretation. Moreover, the phase diagram derived from these simulations contrasts sharply with another conjecture, whereby it is argued that at atmospheric pressure no thermodynamically continuous path exists between low-density amorphous ice and normal water⁶. Here I report the results of a series of molecular dynamics simulations at constant (approximately atmospheric) pressure, which show that both ideas can be reconciled by relocating the critical point to negative pressures. The resulting phase diagram is not only simpler, but also accounts for the transition of water from a fragile to a strong liquid in the supercooled region^{7,8}.

Molecular dynamics simulations were performed at a fixed pressure of 0.1 MPa at several temperatures, using Nosé–Anderson's constant temperature–pressure method^{9–11}. The temperatures, T , were set to 298, 273, 255, 233, 213 and 193 K. The intermolecular interaction was chosen to be TIP4P¹², whose accuracy at lower temperatures was examined in detail by Poole et al.¹³ The simulation time ranged from 1 ns (298 K) to 10 ns (213 and 193 K), significantly longer than previous simulations. This longer run, together with the application of a different method of calculation may lead to different phase behaviours. The number of

molecules is 216. Results of preliminary simulations with a larger number of molecules do not change our conclusion. We selected 500–5,000 configurations each separated by 25 (at higher temperatures) to 2,000 (at lower temperatures) time steps, with a step size of 5×10^{-16} s. These configurations are called instantaneous (I-) structures, which are employed to calculate various properties to be compared directly with experiments, and used to generate quenched (Q-) structures¹⁴, each of which is a corresponding local minimum-energy structure in multi-dimensional energy surface and is free from thermal excitations.

In Fig. 1 we show the temperature dependence of the potential energies, each averaged over a block composed of the above-mentioned 500 successive configurations of both I-structure and Q-structure. The plotted energies of the I-structure have had subtracted from them the harmonic part of the potential, which is $3RT$ where R is the gas constant. The potential-energy surface of a fragile liquid involves numerous multi-level potential energy wells, and what kind of potential energy wells the trajectory can cover depends on the temperature. In contrast, a strong liquid has Arrhenius-type temperature dependence of the viscosity and is considered to have a single excitation process with a regular potential surface¹⁵. The potential energy in the Q-structure decreases significantly with decreasing temperature in the range 298 to 233 K. This leads us to conclude that water in this temperature range is a fragile liquid. A measurement of viscosity supports this view^{7,8}.

More interesting are the large decreases in potential energy of both I- and Q-structures around 213 K. These decreases suggest that, although liquid water at room temperature has a continuous path to a supercooled state down to 233 K, it undergoes a transition around 213 K to another state, whose energy is substantially (1.5 kJ mol^{-1}) lower than that above 213 K, even after removing thermal excitation (open circles in Fig. 1). The potential energy of the Q-structure at 193 K is almost the same as the energy at 213 K. The fact that the potential energy is insensitive to the temperature change implies that water below 213 K (in the TIP4P model) is a strong liquid. It should be noted that this lower-energy phase below 213 K has an energy higher by 1 kJ mol^{-1} than that of proton-disordered cubic ice; the mean potential over 100 Q-

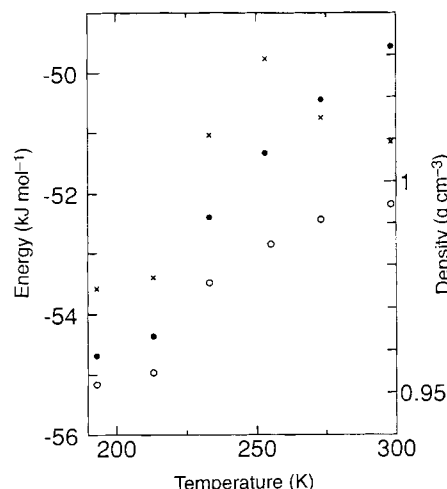


FIG. 1 Potential energy (left-hand ordinate) of instantaneous (I; filled circles) and quenched (Q; open circles) structures of water at various temperatures. Also shown (crosses; right-hand ordinate) is the density of water as a function of temperature. Harmonic energy, $3RT$, is subtracted from the potential energy for I-structures. Individual points are the potential energies or the densities averaged over 500 configurations. The minimum-energy value among 10 blocks at 213 K is plotted to avoid mixing of lower- and higher-energy states owing to a large fluctuation (see text); the chosen block belongs to the lower-energy state, which continues for longer than 0.5 ns.

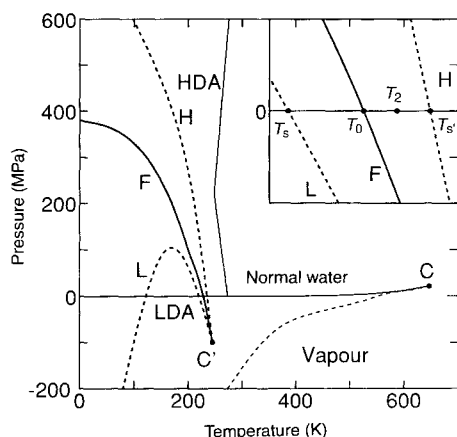


FIG. 2 Schematic phase diagram for water. Those lines concerning LDA–HDA phases are only approximate; the scale of the ordinate axis, pressure, is uncertain. Thin solid lines are coexistence lines of stable phases, which separate liquid water from vapour and ice from liquid water; a break in the ice/liquid–water line is due to the transition from ice Ih to ice III. Thin dashed line is the liquid spinodal. Boundaries of metastable states are shown by a heavy solid line (F) for the coexistence line, and by heavy dashed lines (H and L) for the LDA → HDA spinodal (above which LDA becomes unstable) and for the HDA → LDA spinodal (below which HDA becomes unstable). The gas–liquid critical point is shown by C. The second critical point is denoted by C'. The phase lines for metastable states do not interfere with stable states. A region where the spinodal lines cross the T-axis (that is, $p = 0$) is magnified in the inset. The inequality, $T_s > T_2$, may not hold.

structures of proton-disordered cubic ice in the TIP4P model is $-55.9 \text{ kJ mol}^{-1}$. If this lower-energy phase is identified with low-density amorphous ice (LDA), the calculated energy difference is consistent with the measured heat release at 150 K, 1.3 kJ mol^{-1} (ref. 16).

We propose a new phase diagram of metastable water (Fig. 2) which is, at first glance, similar to that proposed by Stanley *et al.*^{3,4,17} and others^{18,19}. The spinodal line, the limit of stability of high-density amorphous ice (HDA), emerges from this critical point (C'). Both LDA and HDA ices are liquid phases above the glass transition temperature of 136 K (refs 6, 7). As the temperature decreases, the HDA → LDA spinodal line first increases, reaching a maximum pressure point, and then goes down to the negative pressure region again. Other spinodal and coexistence lines probably increase monotonically. A transition from HDA to LDA at a low temperature such as 77 K is unlikely to occur, as is suggested by theoretical calculations²⁰ as well as experiments^{21,22}.

A main difference from the previous study^{3,4,17–19} is the locus of the second critical point (C'), which is located in the negative pressure region. At atmospheric pressure, HDA turns to LDA when the temperature decreases to T_s , the intersection with the HDA → LDA spinodal line. This is consistent with our observation that a substantial decrease in the density is accompanied by the transition from HDA to LDA as shown in Fig. 1. Thus, all the divergences of thermodynamic properties in supercooled water are ascribed to the spinodal instability at T_s , below which HDA becomes unstable. This view is similar to the 'stability-limit conjecture'²² which suggests that anomalies arise from a spinodal instability originating at the gas–liquid critical point, shown as C in Fig. 2. Although an instability at T_s is not due to re-entry of the gas–liquid spinodal line from the negative pressure region (proposed in the stability-limit conjecture) but to the HDA → LDA spinodal line in our picture, the idea that divergences arise from the limit of stability is common to both pictures. It is reasonable that normal water, as designated by Speedy⁶, is identified with water at temperature above the critical point. In our phase diagram, LDA is a different phase from water in the supercooled state, so that large fluctuations associated with the critical point are not observed in the high-pressure region, which is consistent

with the fact that water ceases to exhibit singular behaviour above 0.2 GPa.

There must exist a critical point at around 213 K and at a pressure lower than atmospheric. This is because a large density fluctuation, 0.03 g cm^{-3} at 213 K compared with 0.01 g cm^{-3} at 233 K, is attributed to the transition between two distinct phases. This situation is the same as the density of system in isothermal-isobaric condition near the gas–liquid critical point; the density fluctuates between gas and liquid phases. In accordance with the large density fluctuation, the energy difference between the maximum and minimum energies at 213 K among the blocks (each composed of 500 configurations) amounts to 0.5 kJ mol^{-1} , which is much larger than 0.1 kJ mol^{-1} at 233 K and 0.05 kJ mol^{-1} at 193 K. It is well known that, although pressure is correct in TIP4P, the temperature associated with this potential should be shifted upward by 10–20 K in order to agree with experiment³. That the divergences occur experimentally at 228 K is thus in correspondence with the discontinuity at about 213 K in our molecular dynamics simulation.

As water is cooled, it encounters the LDA → HDA spinodal line at T_s and the coexistence line at T_0 . No transition occurs until the HDA → LDA spinodal line is reached at T_s . However, crystallization intervenes at T_2 , and we can observe only the 'symptom' of critical fluctuations in thermodynamic response functions. When heated LDA should become HDA (normal water) after crossing T_s . This transition is again not observed experimentally because LDA becomes cubic ice at T_1 ($\sim 160 \text{ K}$). Speedy⁶ pointed out that the entropy difference between cubic ice and LDA must be smaller (to justify this scenario) than the estimated value based on the random network model²⁴. However, its application here relies on the existence of a continuous path from normal water to LDA. This is not the case in our picture, and the entropy difference can be smaller than that calculated from the random network model. Therefore, our phase diagram may not suffer from any apparent thermodynamic inconsistency.

In Fig. 3 we show the distributions of pair interaction energy for Q-structures. Only pairs of water molecules separated by less than 3.5 Å are taken into consideration. It is clear, from a comparison between those at 298 K and those at 233 K, that a decrease in temperature reduces the number of high-energy (defect) pairs while the distribution of strongly hydrogen-bonded (stable) pairs remains almost unchanged. A further decrease in temperature by only 20 K, however, changes the distribution drastically. Most of the defects suddenly disappear, resulting in formation of stable hydrogen bonds. The radial distribution function for LDA at 213 K (and 193 K) is quite different from that of ice. The density of states for intermolecular vibrational motions calculated from Q-structures are given in Fig. 4. Difference between those at 298 K

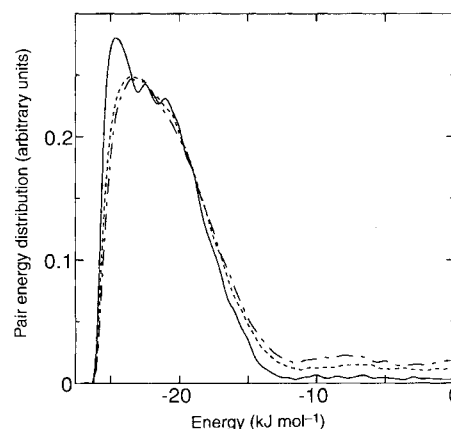


FIG. 3 Distributions of pair interaction energy for Q-structures at temperatures of 213 K (solid line), 233 K (dotted line) and 298 K (dash-dot line).

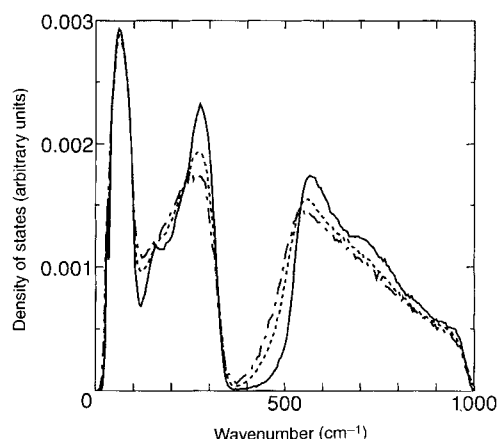


FIG. 4 Density of states for intermolecular vibrational motions at temperatures of 213 K (solid line), 233 K (dotted line) and 298 K (dash-dot line).

and 233 K is insignificant, whereas the difference is large between those at 233 K and 213 K. Lowering the temperature by only 20 K reduces number of modes between 300 to 500 cm^{-1} , which are mixed modes composed of translations and rotations of individual molecules and lead to a neighbouring Q-structure with small excitation energy²⁵.

A simple extrapolation of the potential energy for I-structure above 233 K intersects with the energy of cubic ice in its Q-structure at around 160 K. Unless there is a break of the potential energy plot in Fig. 2 between 233 K and 160 K, the energy of the liquid state becomes lower than that of ice. In order to preclude this Kauzmann's paradox^{26,27}, water must undergo a fragile–strong type transition before vitrification as pointed out by Angell⁷. The transition observed around 213 K is not a glass transition as individual molecules even at 193 K still move more than the molecular separation, 2.8 Å, for 10 ns and the mean square displacement is linear in time, which ensures that the diffusion is normal. The self-diffusion coefficients at 233, 213 and 193 K are 2×10^{-6} , 9×10^{-7} and $6 \times 10^{-7} \text{ cm}^2 \text{ s}^{-1}$, respectively. This indicates that water at 193 K is a liquid phase, termed water II.⁶

Our preliminary simulation results at -0.2 GPa show that those properties given in Figs 3 and 4 have a much smaller temperature dependence than those at atmospheric pressure. It is therefore reasonable to locate the second critical point between 0 and -0.2 GPa , thereby avoiding the introduction of a new phase. Our phase diagram should be tested against direct calculations of the free energy of LDA and HDA. We also consider it important that a theoretical investigation be made of the temperature dependence of the potential surfaces and dynamics of our model, to delineate the mechanism of the transition suggested here. □

Received 11 August 1995; accepted 27 February 1996.

- Speedy, R. J. & Angell, C. A. *J. chem. Phys.* **65**, 851–858 (1976).
- Angell, C. A., Oguni, M. & Sichiya, W. *J. phys. Chem.* **86**, 998–1002 (1982).
- Poole, P. H., Sciortino, F., Essmann, U. & Stanley, H. E. *Nature* **360**, 324–328 (1992).
- Poole, P. H., Sciortino, F., Essmann, U. & Stanley, H. E. *Phys. Rev. E* **48**, 3799–3817 (1993).
- Eisenberg, D. & Kauzmann, W. *The Structure and Properties of Water* (Oxford Univ. Press, London, 1969).
- Speedy, R. J. *J. phys. Chem.* **96**, 2322–2325 (1992).
- Angell, C. A. *J. phys. Chem.* **97**, 6339–6341 (1993).
- Angell, C. A. in *Hydrogen Bond Networks* (eds Bellissent-Funel, M.-C. & Dore, J. C.) 3–22 (Kluwer, Dordrecht, 1994).
- Nosé, S. *Molec. Phys.* **52**, 255–268 (1984).
- Nosé, S. *J. chem. Phys.* **81**, 511–519 (1984).
- Andersen, H. C. *J. chem. Phys.* **72**, 2384–2393 (1980).
- Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. *J. chem. Phys.* **79**, 926–935 (1983).
- Poole, P. H., Essmann, U., Sciortino, F. & Stanley, H. E. *Phys. Rev. E* **48**, 4605–4610 (1993).
- Stilling, F. H. & Weber, T. A. *J. phys. Chem.* **87**, 2833–2840 (1983).

- Stilling, F. H. *J. chem. Phys.* **88**, 7818–7825 (1988).
- Hallbrucker, A. & Mayer, E. *J. phys. Chem.* **91**, 503–505 (1987).
- Stanley, H. E. *et al. Physica A* **205**, 122–139 (1994).
- Ponyatovskii, E. G., Sinitsyn, V. V. & Pozdnyakova, T. A. *JETP Lett.* **60**, 360–364 (1994).
- Bellissent-Funel, M.-C. & Bosio, L. *J. chem. Phys.* **102**, 3727–3735 (1995).
- Poole, P. H., Sciortino, F., Grande, T., Stanley, H. E. & Angell, C. A. *Phys. Rev. Lett.* **73**, 1632–1635 (1994).
- Mishima, O., Colvert, L. D. & Whalley, E. *Nature* **310**, 393–394 (1984).
- Mishima, O., Colvert, L. D. & Whalley, E. *Nature* **314**, 76–78 (1985).
- Speedy, R. J. *J. phys. Chem.* **86**, 982–991 (1982).
- Sceats, M. S. & Rice, S. A. *J. chem. Phys.* **72**, 3260–3262 (1980).
- Ohmine, I. & Tanaka, H. *J. chem. Phys.* **93**, 8138–8147 (1990).
- Kauzmann, W. *Chem. Rev.* **43**, 219–256 (1948).
- Stanley, H. E. & Teixeira, J. *J. chem. Phys.* **73**, 3404–3422 (1980).

ACKNOWLEDGEMENTS. I thank I. Ohmine, M. Sasai, H. E. Stanley, R. Yamamoto and K. Koga for helpful discussions, and I. Okabe for providing cubic ice structures. This work was supported by the Japan Ministry of Education and by the Computer Center of IMS.

Palaeobotanical evidence for a warm Cretaceous Arctic Ocean

Alexei B. Herman* & Robert A. Spicer†

* Geological Institute of the Russian Academy of Sciences, Moscow 109017, Russia

† Department of Earth Sciences, The Open University, Milton Keynes MK76AA, UK

THE Cretaceous period was a time of global warmth^{1–4}. Mid-Cretaceous equatorial temperatures were similar to today's⁵, but the equator-to-pole temperature gradient is the subject of some controversy^{5–7}. Although it is unlikely that the poles were ice-free^{8–10}, fossil evidence^{3–5,11,12} indicates that near-polar temperatures were much higher than they are today. Little is known, moreover, about oceanic poleward heat transport, and this makes it hard to model the Cretaceous climate or to evaluate the extent to which it provides an analogue for a 'greenhouse' world warmed by increased atmospheric CO_2 alone. Here we use relationships between leaf physiognomy (such as shape and size) and modern climate to determine Cretaceous climate conditions in the Arctic region from fossil leaves. We find that the Arctic Ocean was relatively warm, remaining above 0°C even during the winter months. This implies that there was significant poleward heat transport during all seasons.

The Cretaceous Arctic Ocean had only shallow connections to lower latitudes, and so if it was warmed by significant oceanic heat transport, the susceptibility of the more open Arctic Ocean to warming becomes a particularly pressing concern. But attempts to model the Cretaceous climate using atmospheric general circulation models generally treat the oceans in a simplistic manner, prescribing the sea surface latitudinal thermal gradient¹³ rather than determining it from the model. This makes it important to have accurate knowledge of this gradient in order to model the Cretaceous palaeoenvironment.

One of the most useful, sensitive and quantitative sources of non-marine palaeoenvironmental data is the morphology of fossil leaves. Today, particular plant morphologies are found repeatedly in a given climate regime. This physiognomic expression is robust in space and time because it is controlled primarily by the physical laws of gas diffusion, fluid transport and evaporation. It is the basis for taxon-free vegetation mapping^{14–16} and has proved particularly successful when applied to leaves^{17,18}. Early foliar physiognomic studies used the ratio of toothed-margined to smooth-margined woody dicot leaf taxa to derive average annual air temperatures¹⁷. Newer techniques^{19,20} employ a multivariate statistical analysis (correspondence analysis, CA^{21} , or canonical correspondence analysis, $\text{CANOCO}^{22,23}$) to identify and calibrate correlations between numerous leaf characters and climate variables (here we use 31 of the former and 8 of the latter). CANOCO is a direct ordination method, and we use it here to order site, leaf character

and environmental data simultaneously in multidimensional space: sites are ordered by their character scores, and characters by their distribution among the sites. The sites are therefore arranged relative to one another in multidimensional space using the physiognomic characters of the vegetation at that site (Fig. 1a); environmental data are not used to position the sites. CANOCO explicitly positions the environmental vectors within this physiognomically defined vegetation space (Fig. 1b) and is therefore preferable to CA, which relies on subjective alignment of the environmental vectors. The environmental vectors are calibrated using modern sites with known climates (see, for example, Fig. 1c, d) and palaeoclimate data are derived from the fossil site vector scores.

We modified Wolfe's¹⁹ method by using CANOCO to analyse a matrix of 103 modern vegetation samples, for which the climate was known, each scored for 31 foliar characters which are correlated to 8 climate variables¹⁹. Our analysis differs by including vegetation from seasonally arid regions of Mexico, and two additional leaf size classes (nanophyll and mesophyll III). Fossil data, entered into the analysis as passive²² samples, plotted well clear of the Wolfe's 'sub-alpine nest'¹⁹. Alpine sites were subsequently excluded from the reference data.

Two fossil floras of Turonian age (91–88 Myr; one from Kamchatka and one from Novaya Sibir' island), and two Coniacian (88–86 Myr; one from the North Slope of Alaska and the other from Kamchatka) (Fig. 2) were scored for as many of the 31 characters as were preserved. As many specimens as possible within a given leaf taxon were scored for character variation; missing characters were given a null score. Only four taxa out of the 94 analysed lacked margin data, which are important in deriving temperatures, while size data, important in deriving

precipitation estimates, were unreliable in only two taxa. These details will be described elsewhere²⁴.

The results are shown in Table 1. The most northerly samples for each time horizon exhibited the highest values for mean annual temperature (MAT), cold month mean temperature (CMMT) and all precipitation data. The MAT values are consistent with earlier palaeobotanical estimates using less sophisticated techniques^{4,12}, and the high precipitation estimates are consistent with the abundance of Arctic Cretaceous coals^{25,26}. Warm month mean temperature (WMMTs) estimates are uniformly between 17.8 and 20.0 °C, but CMMTs range from –3.8 to 5.7 °C. In sites closest to the Arctic Ocean, CMMTs are at or above freezing in spite of winter darkness lasting between 1.5 (at 75° N) and 3 months (at 82° N)²⁷. We interpret these results in terms of a warm Arctic Ocean, with significant poleward marine heat transport. A warm Arctic Ocean would have damped seasonal temperature oscillations, contributing little to summer warmth but maintaining winter temperatures above freezing. High evaporation would have formed extensive cloud cover, which would have helped maintain warm air temperatures, and contributed to high precipitation.

The Turonian and Coniacian stages have been reconstructed as being respectively a cool and a warmer interval in the early Late Cretaceous^{4,28}, both were times of high relative sea levels and transgression in the Western Interior Seaway (WIS)²⁹. The Alaskan Coniacian CMMT of 5.7 °C may have been due to the site's proximity to warm water entering the Arctic Ocean from the equatorial Tethys Ocean via the WIS, with a return path southwards through the proto North Atlantic. To maintain this air temperature throughout the winter darkness, the adjacent sea surface temperature (SST) must have been at least 6 °C and

FIG. 1 Some results of the multivariate analysis of foliar physiognomic data. a, Axis 1/axis 2 plot for vegetational sites coded for mean annual temperature, MAT; low-MAT sites plot to the left, warm sites to the right. This trend is represented by the MAT vector (9) in b. Other climate trends are shown by the vectors in b. Fossil sites are: AC, North Slope Alaska Coniacian; KC, Kamchatka Coniacian; KT, Kamchatka Turonian; NT, Novaya Sibir' Island Turonian. b, Environmental vectors in axis 1/axis 2 space. Labels as follows: 1, precipitation during the three driest months; 2, mean annual precipitation; 3, mean monthly growing season precipitation; 4, mean growing season precipitation; 5, cold month mean temperature, CMMT; 6, MAT; 7, length of the growing season; 8, warm month mean temperature, WMMT. c, d, Examples of how climate variables are derived from the environmental vector scores by means of a second-order polynomial regression. All climate variables given in Table 1 are similarly derived. c, Plot of the MAT vector scores against observed MATs with fossil scores as indicated. d, Plot of the cold month mean vector score against observed CMMT. The cumulative percentage variation accounted for by axes 1, 2 and 3 (3 not shown) were 50.5%, 60.9% and 64.3% respectively. Environmental vectors lie close to, but are not coincident with, the major axes of variation, therefore to use the major axes of variation for environmental determination is not ideal. Instead the environmental vectors themselves were calibrated to obtain palaeoclimatic data (for example, Fig. 2c and d). Environmental vector scores are the distances along a vector at which a site with a known climate projects normally from axis 1/axis 2 space. Environmental vector scores were then plotted against the observed climate variable for that site and a regression curve fitted. Vector scores for fossil sites are then projected on to the regression curve to derive a predicted palaeoclimate variable (for example, c and d). R^2 values for the fit of the curve to the data are given in Table 1, as are the standard deviations of the residuals about the mean for each curve. The cumulative percentage variance in species–environment relationships for axes 1, 2 and 3 were 76.0%, 91.6% and 96.7% respectively.

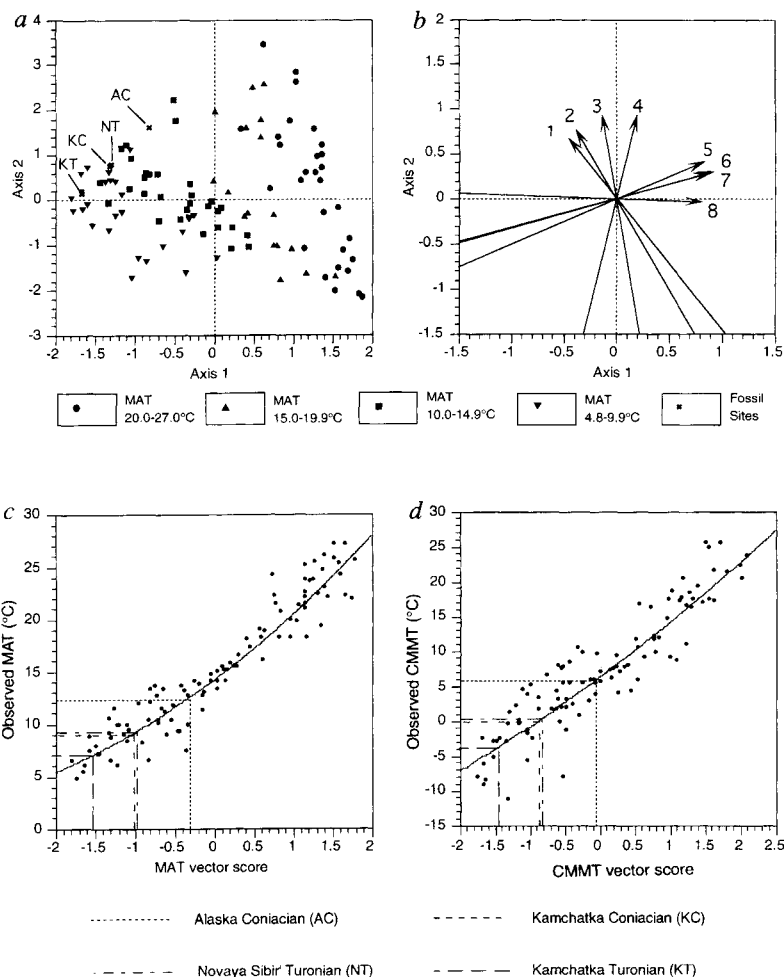


TABLE 1 Climate data for the Turonian and Coniacian floras of northern Alaska and Arctic and northeastern Russia

Locality	North Slope	Kamchatka	Kamchatka	Novaya Sibir'	Regression line	R^2	S.d. of residuals
Age	Coniacian	Coniacian	Turonian	Turonian			
Palaeolatitude	75° N	72° N	72° N	82° N			
Geological/ phytostratigraphic unit and formation	Tuluvak Tongue, Chandler Fmtn	Kaivayam Horizon, Valizhgen Fmtn	Penzhina Horizon, Valizhgen Fmtn	Not differentiated. Derevyannye Gory Fmtn			
Number of leaf taxa analysed	25	28	21	20			
Data completeness statistic*	0.87	1.31	1.47	0.96			
Mean annual temperature (°C)	12.5	9.0	6.9	9.0	$y = 14.104 - 5.632x + 0.633x^2$	0.913	1.8
Warm month mean temp. (°C)	20.0	18.6	17.8	18.5	$y = 22.46 + 4.01x + 0.684x^2$	0.663	3.1
Cold month mean temp. (°C)	5.7	00	-3.8	0.0	$y = 6.18 + 7.456x + 0.382x^2$	0.845	3.3
Length of growing season† (months)	7.4	5.7	4.7	5.7	$y = 8.116 - 2.515x + 0.192x^2$	0.847	1.1
Mean annual precipitation (mm)	1,638	1,436	1,348	1,455	$y = 922.891 + 4.6166x + 46.427x^2$	0.538	430
Mean growing season precipitation (mm)	1,287	827	598	852	$y = 419.146 + 306.982x + 115.946x^2$	0.687	280
Mean monthly growing season precipitation (mm)	122	86	67	88	$y = 58.114 + 28.527x + 6.177x^2$	0.716	23
Three consecutive dry month precipitation (mm)	271	205	170	211	$y = 76.888 + 72.269x + 22.39x^2$	0.585	70

The sixth column shows the equation of the second-order polynomial regression line for environmental vector scores versus observed environmental data, fitted using Statview 4.1. In all cases $P < 0.0001$. The rightmost column gives the standard deviation of residuals about the regression line. Palaeolatitudes are derived from Smith *et al.*³⁶ and stratigraphic data were derived from published sources^{4,31,32,34,35}. With the exception of the Novaya Sibir' flora, all material was directly examined and scored from photographic enlargements of the leaves. The Novaya Sibir' flora was scored from a published work³¹. Assuming an obliquity similar to that of the present, the growing season durations are the maximum possible if only light is limiting to plant growth, not temperature as in today's climate. Previous work¹² has shown this assumption to be realistic.

* The data completeness statistic is a measure of the completeness of the data matrix and is calculated as follows: $(F - M)/P$ where F is the actual number of data matrix cells filled, M is the minimum number of data matrix cells missing, and P is the minimum number of data matrix cells that should be filled if all character states were to be scored.

† Growing season—all months with a mean temperature of at least 10 °C.

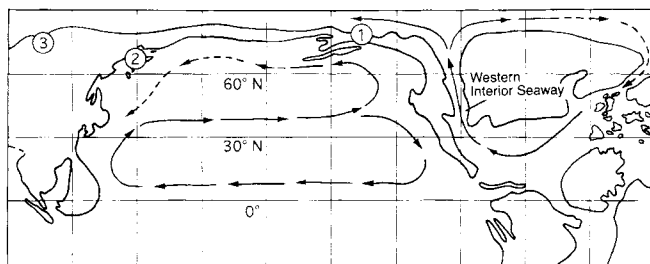


FIG. 2 Locations of the four fossil floras used in this study plotted on a map of Coniacian coastlines derived from Funnel³⁰. 1, Coniacian site, North Slope, Alaska (USGS Locality No. 11606); 2, Turonian and Coniacian sites from Kamchatka (Locality Nos 3/3-4/9, 829, 2-1, 2-2¹¹); 3, Turonian site from Novaya Sibir' Island³¹. Stratigraphic positions of all but the Novaya Sibir' flora were based on fossils from marine interbeds^{4,11,32-35}, and represent sites near Cretaceous sea level. Dating of the Novaya Sibir' deposits is based on palynological and megafloral correlation with floras from independently dated beds in northeastern Russia³¹. Each flora was collected from a range of lithologies and sedimentary environments thereby minimising the taphonomic bias associated with any single depositional system or ecophysiological constraints associated with any single plant community. Ocean circulation, inferred from the data given here, is shown by arrows. Solid arrows represent warm currents; dashed arrows, cool currents.

possibly 8 °C. The Novaya Sibir' CMMT may have been cooler because of its greater distance along the current path from the WIS, a cooler climate in the Turonian, and the higher palaeolatitude of the site. The nearby Arctic Ocean SST cannot have been below 0 °C, or the adjacent land would have cooled to below freezing during the protracted winter darkness. The relative winter coolness of the Kamchatkan sites argues against atmospheric poleward heat transport because warm north-flowing air would have been recorded by the Kamchatka vegetation and expressed as a high predicted CMMT. That such an effect is not apparent suggests that heat was transported in the more channelled marine medium.

Our results are consistent internally and with qualitative interpretations of sediments, and do not violate independent environmental constraints such as the polar light regime. The predicted lengths of the growing seasons (thermally defined in the analysis) coincide with the duration of summer light (near-present-day obliquity is assumed¹²), suggesting that duration of sunlight, not temperature, was the primary growth-limiting factor. The results suggest that the Cretaceous mean annual surface latitudinal temperature gradient was small at high latitudes, as assumed in a recent modelling study⁷. □

Received 20 September 1995; accepted 29 February 1996.

1. Chumakov, N. M. *et al.* *Stratigr. geol. Correlation* **3**, 241–260 (1995).
2. Spicer, R. A. & Chapman, J. L. *Trends Ecol. Evol.* **5**, 279–284 (1990).

3. Vakhrameev, V. A. *Jurassic and Cretaceous Floras and Climates of the Earth* (Cambridge Univ. Press, 1991).
4. Herman, A. B. in *Cenozoic Plants and Climate of the Arctic* (eds Boulter, M. C. & Fisher, H. C.) 127–151 (NATO ASI Ser., Springer, Heidelberg, 1994).
5. Sellwood, B. W., Price, G. D. & Valdes, P. J. *Nature* **370**, 453–455 (1994).
6. Spicer, R. A. & Corfield, R. M. *Geol. Mag.* **129**, 168–180 (1992).
7. Barron, E. J., Fawcett, P. J., Peterson, W. H., Pollard, D. & Thompson, S. L. *Paleoceanography* **10**, 953–962 (1995).
8. Barron, E. J., Thompson, S. L. & Schneider, S. H. *Science* **212**, 501–508 (1981).
9. Francis, J. E. & Frakes, L. A. *Sedimentol. Rev.* **1**, 17–30 (1993).
10. Spicer, R. A. & Parrish, J. T. *Geology* **14**, 703–706 (1986).
11. Herman, A. B. in *Stratigrafiya i flora melovykh otlozhenii Severo-Zapadnoi Kamchatki (Stratigraphy and Flora of the Cretaceous Deposits of Northwestern Kamchatka)* (eds Herman, A. B. & Lebedev, E. L.) 5–141 (Nauka, Moscow, 1991).
12. Parrish, J. T. & Spicer, R. A. *Geology* **16**, 22–25 (1988).
13. Valdes, P. J. & Sellwood, B. W. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **95**, 47–72 (1992).
14. Box, E. O. *Macroclimate and Plant Forms: An Introduction to Predictive Modelling in Phytogeography* (Junk, The Hague, 1981).
15. Holdridge, L. R. *Science* **105**, 367–368 (1947).
16. Walter, H. *Der ökologische Systeme der Kontinente (Biogeosphäre)* (Fischer, Stuttgart, 1976).
17. Bailey, I. W. & Sinnott, E. W. *Science* **41**, 831–834 (1915).
18. Wolfe, J. A. *Prof. Pap. U.S. geol. Surv.* **1106**, 37 (1979).
19. Wolfe, J. A. *Bull. U.S. geol. Surv.* **2040**, 1–71 (1993).
20. Wolfe, J. A. *Rev. Earth planet. Sci. Lett.* **23**, 119–142 (1995).
21. Hill, M. O. *Appl. Statist.* **23**(3), 340–354 (1974).
22. ter Braak, C. J. *Ecology* **767**, 1167–1179 (1986).
23. ter Braak, C. J. F. *CANOCO—A FORTRAN Program for Canonical Correspondence Ordination* (Micropower Power, Ithaca, New York, 1987–1992).
24. Herman, A. B. & Spicer, R. A. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* (submitted).
25. Krassilov, V. A. *Spec. Pap. geol. Soc. Am.* **267**, 263–267 (1992).
26. Spicer, R. A., Parrish, J. T. & Grant, P. R. *Spec. Pap. geol. Soc. Am.* **267**, 177–192 (1992).
27. *Handbook, Polar Regions Atlas 1–66* (CIA, National Foreign Assessment Center, Washington DC, 1978).
28. Zakharov, Y. D., Grabovskaya, V. S. & Kalishevich, T. G. in *Sistematika i evoliutsiya bespozvochnykh Dal'nego Vostoka* (in Russian) 41–90 (DVNTS AN SSSR, Vladivostok, 1984).
29. Hills, L. V., Braunberger, W. F., Núñez-Betelu, L. K. & Hall, R. L. *Can. J. Earth Sci.* **31**, 733–736 (1994).
30. Funnell, B. M. in *Cretaceous Resources Event and Rhythms* (eds Ginsburg, R. N. & Beaudoin, B.) 221–235 (Kluwer Academic, Netherlands, 1990).
31. Sveshnikova, I. N. & Budantsev, L. Y. *Iskopaemye flory Arktyki, I (Fossil Floras of the Arctic, I)* (in Russian) (Nauka, Leningrad, 1969).
32. Brosgé, W. P. & Whittington, C. L. *Prof. Pap. U.S. geol. Surv.* **303-H**, 501–638 (1966).
33. Herman, A. B. *Stratigr. geological Correlation* **2**, 365–378 (1994).
34. Pergament, M. A. *Stratigrafiya i inostrany verkhnego mela severnogo polushariya (Stratigraphy of Inoceramids of Upper Cretaceous of the Northern Hemisphere)* (in Russian) 1–214 (Nauka, Moscow, 1978).
35. Poikhalainen, B. P. *Late Senonian Heteromorphic Ammonites of the Anadyr-Koryak Region (Stratigraphy and Palaeontology of the Palaeontology of the Palaeozoic and Mesozoic Deposits of the North-East USSR)* (in Russian) 1–195–205 (Nedra, Moscow, 1984).
36. Smith, A. C., Hurley, A. M. & Briden, J. C. *Phanerozoic Palaeocontinental World Maps 1–102* (Cambridge Univ. Press, 1981).

ACKNOWLEDGEMENTS. We thank J. A. Wolfe for his advice and inspiration, and P. Valdes for discussions on ocean heat transport. This work was supported by the Royal Society.

Analogies to oceanic behaviour in the continental breakup of the western Woodlark basin

John C. Mutter^{*†}, C. Z. Mutter^{*} & J. Fang^{*†}

^{*} Lamont-Doherty Earth Observatory of Colombia University, Palisades, New York 10964, USA

[†] Department of Geological Sciences, Colombia University, New York, New York 10027, USA

THE Woodlark basin/D'Entrecasteaux Island region off northeast Papua New Guinea (Fig. 1) offers a rare glimpse of the propagation of active rifting and continental breakup¹ into orogenically thickened (and subsequently extended) lithosphere^{2–10}. Rifting, continental breakup, and the subsequent formation of oceanic lithosphere by sea-floor spreading often involve rupture controlled by propagation of a rift tip^{11–20}. We recognize several tectonic elements in the evolving western Woodlark intra-continental system that are close geometric analogues of wholly oceanic propagators, and a kinematic development that can be viewed as an extrapolation of oceanic tectonics. The response to deformational stresses in the final stages of breakup in a thick, relatively hot, and hence weak continental lithosphere appears akin to that of oceanic lithosphere.

In 1992, we obtained the first systematic deep seismic reflection profiling²¹ covering the region from 150 km west (ahead) of the active Woodlark propagator, to more than 200 km east of the tip where sea-floor spreading in the Woodlark basin has been occurring for ~4 Myr (Figs 1, 2). East of the propagator tip (post-breakup) the data describe a strong asymmetry in the emerging conjugate passive margin pair that persists to at least 154° E, suggesting that it has characterized at least the past million years of breakup. The Woodlark rise (northern margin of the basin) is a generally undisturbed feature except for its abrupt southern boundary formed by a set of high-angle normal faults stepping down to the south (Figs 1, 2c, d). This margin exhibits seismic activity in contrast to the southern margin, which is quiet. The Pocklington rise (southern margin of the basin) is characterized by large, high-relief fault-bounded crustal blocks that typically rise to much shallower depths than the Woodlark rise. The northern boundary of the Pocklington rise has a two-part structure (Fig. 1). East of 151° 45' E, the boundary has a segmented^{1,22} but generally arcuate shape with northwest–southeast orientation. West of 151° 45' E, where *bona fide* oceanic crust has yet to form, the

boundary is formed by the southern flank of a deep, east–west oriented depression incorporating discrete basins like North and South valleys²³ at the westernmost extremity. These depressions are the zones of active intra-continental deformation most proximal to the oceanic spreading system.

Moresby seamount (Fig. 1) is a conical edifice located just west and south of the propagating rift tip. This feature is actually an extremely large fault block¹ with a prominent fault plane forming its northern flank, dipping beneath the Woodlark rise (Fig. 2c, d). We have suggested²⁴, as has Taylor *et al.*¹, that Moresby seamount forms the footwall of a major low-angle normal fault²¹ along which the deep crust of the lower Woodlark rise has been exhumed. Its dimensions and relief are similar to those of the individual core zone domes on the D'Entrecasteaux Islands (Fig. 1), and the vergence of major bounding faults on its northern flank is also similar to dome-bounding shear zones on the northern flanks of the islands' core complexes^{3,5}. These attributes allow us to propose that Moresby seamount may be a submerged metamorphic core complex.

West and south of the propagator (Figs 1, 2a, b) our seismic images show that the central Goodenough basin includes a distinct basement high with dimensions that are quite similar to those of Moresby seamount (compare Fig. 2a and b, c and d), and is paired to the north with a depression that is also similar in scale to features such as South valley²³ (immediately north of Normanby Island, Fig. 1). The depression is filled with strata dipping steeply south into the well defined northern flank of the basement high (Fig. 2b). The entire cross-section, including Fergusson Island, appears to be formed of a series of crustal blocks, decreasing in size to the south. In the southern extremity of the basin, sediments dip steeply southward towards Papua New Guinea, providing evidence for a major normal fault along the coastline (Fig. 1), although a coastal fault system itself has not been imaged in our data. Seismicity is generally confined to the southern part of the basin, where the most recent faulting occurs within flat-lying to SW-dipping sediments. Here, active extensional and rotational deformation forms a distinct zone which we call the Goschen deformation zone (GDZ, Figs 1, 3). The GDZ forms a graben (Fig. 3a) that broadens then merges into a prominent half-graben (Fig. 3b) eastwards towards Goschen strait (south of Normanby Island).

These observations lead us to suggest that propagation of continental breakup in the western Woodlark basin may not require a unique formulation, but may be understood by extrapolation from phenomena already well described in purely oceanic settings^{25–27}. The simple, linear boundary of the Woodlark rise defines an outer pseudo-fault (OPF), the Woodlark pseudo-fault (WPF; Fig. 4a) which is analogous to that of an oceanic

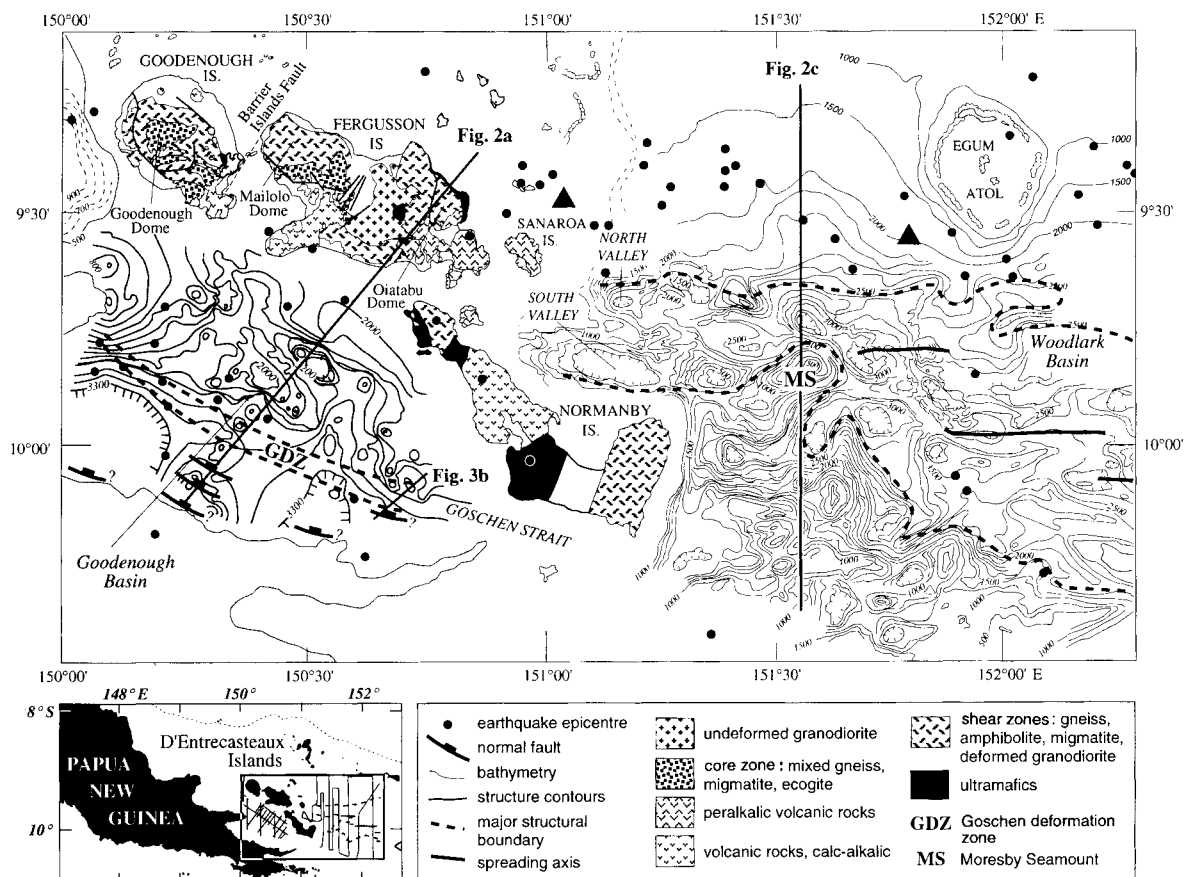


FIG. 1 Large map shows features mapped from seismic imaging combined with geological and other geophysical data^{1-10,17,18} in the Woodlark basin propagating spreading centre/intra-continental extensional province. Locations of Fig. 2 and Fig. 3b profiles also shown. Earthquake locations with local mechanism solutions permitting shallow, north to NNW dipping normal faults²⁸ shown by triangles. Small map shows location of study area and multichannel coverage; the key identifies structural designations and exposed rock types. The D'Entrecasteaux Islands (Goodenough, Fergusson and Normanby) and perhaps the youngest metamorphic core complexes on Earth⁴⁻⁸. Major WNW-trending faults separate the gneissic domes from the underlying ultramafic rocks on the northeastern edges of the domes. The centres of the domes expose massive granodiorite bodies^{3,7} that emerged beginning about 4 Myr ago^{5,6} when westward-directed rift propagation progressively ruptured thickened lithosphere of Papua New Guinea, creating

propagator (Fig. 4b). Similarly, the arcuate-shaped boundary between the Woodlark basin and the Pocklington rise is the expression of an analogous inner pseudo-fault (IPF), the Bonvouloir pseudo-fault (BPF). Like wholly oceanic inner pseudo-faults (Fig. 4b), the detailed topography of the BPF is more complex than the general structural trends represented by the solid lines (Fig. 4a). East of 151° 30' E (Fig. 4a), the WPF and BPF define the boundary between new oceanic lithosphere and the existing continental lithosphere. West of 151° 30' E, the WPF and BPF can be projected to form opposite flanks of a region of roughly east-west oriented depressions that include North and South valleys (Fig. 1). Immediately northwest of North valley is the location of the larger of the earthquakes that have nodal plane solutions consistent with low-angle normal faulting²⁸. The disturbed topology of the westernmost Pocklington rise, South valley and part of Normanby Island (Fig. 1) occupies a position that is the geometric analogue of an oceanic migrating extensional relay zone (MERZ; Fig. 4b) and we propose that this region also formed as a zone of lithospheric transfer (MERZ; Fig. 4a). The GDZ in the southern Goodenough basin (Fig. 4a) lies in the analogous location to the doomed rift (DR; Fig. 4b) of an oceanic

the Woodlark and Pocklington rises as a conjugate passive margin pair. Propagation is evidenced by a westward narrowing of the basin and progressive termination of older sea-floor magnetic lineations⁴. Individual spreading segments exhibit a distinct east-west orientation. The propagating rift tip presently lies near 151° 30' E. An intermediate-sized ($M_w = 6.8$), shallow (2.6 km) earthquake located north of Sanaroa Island (triangle) permits a normal fault mechanism with one plane having an unusually shallow dip of 10–25° dipping north to NNW²⁸. Present-day tectonism is focused in the deep corridor just west of the propagating rift tip and in the Goodenough deformation zone (GDZ; southern Goodenough basin) that projects eastwards into Goschen straits south of Normanby Island. Structural features within the GDZ reflect both extensional and strike-slip motion (folds and 'flower' structures are present) suggesting an overall trans-tensional mode of deformation.

propagator. The GDZ, however, occurs in a wholly intra-continental setting.

Under the proposed geometric analogies, the Solomon/Indo-Australian plate boundary is defined by the Woodlark spreading centre that has propagated westwards to almost 151° 30' E (Figs 1, 3a). West of the propagating rift (PR) tip, to about 151° E, extension is occurring in the east-west corridor incorporating North and South valleys and is perhaps associated with low-angle normal faulting in the region of Sanaroa island. West of 151° E the strain associated with plate separation is at present accommodated in the GDZ, a feature that parallels the westernmost corridor of extensional deformation (North and South valleys) associated with the oceanic propagator. A near 'overlap' situation is developed that mimics the geometry of a purely oceanic propagator (Fig. 4). Kinematic evolution of the system involves propagation of the spreading centre and continuous redefinition of the intra-continental deformation zones, presently represented by the GDZ and North and South valleys, that progressively concentrates deformation towards the propagator. In Goschen strait, deformation instantaneously defines a new plate boundary that must accommodate approximately the

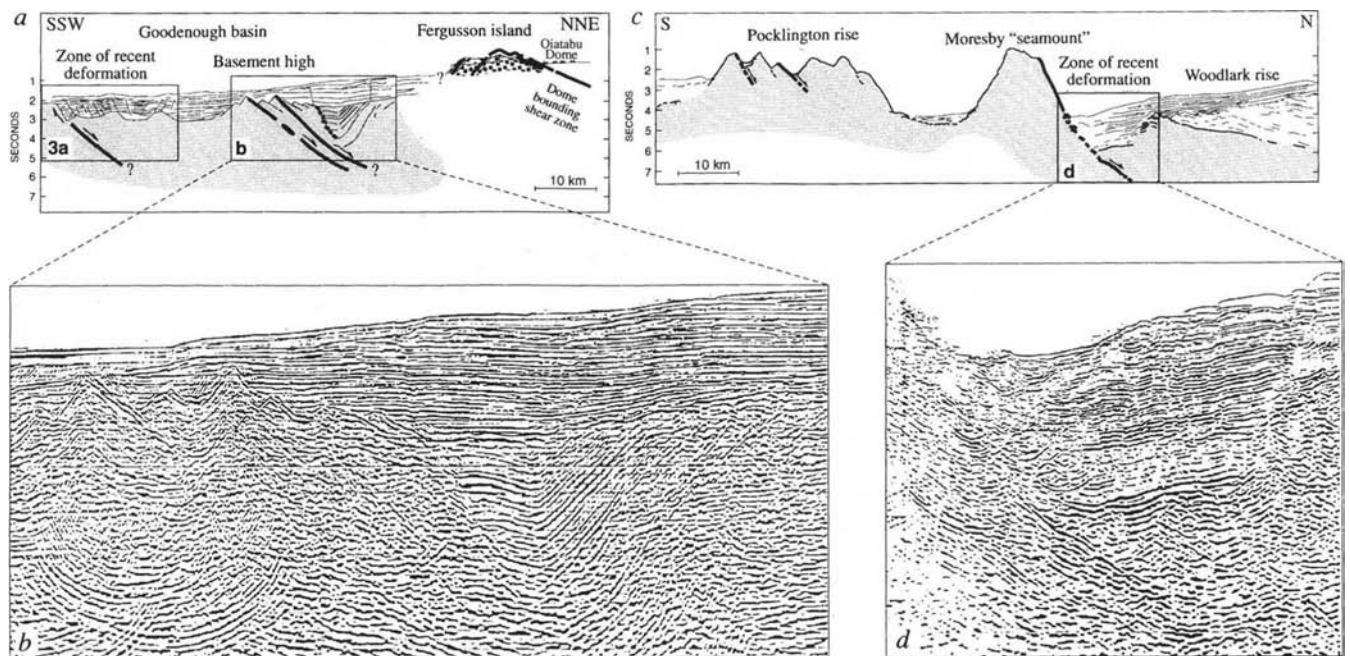


FIG. 2 Line drawings (a, c) and selected portions of seismic images (b, d) obtained as part of the seismic study (locations shown in Fig. 1). Illustration a is constructed using seismic profiles 1194 and 1198 in the Goodenough basin together with geological information from Fergusson Island^{18,20} (see Fig. 1 for identification of exposed rock type), and identifies locations of seismic profiles shown in b and Fig. 3a. Image b is a profile from the central basin, showing a distinct basement high paired to the north with a depression filled with strata dipping steeply to the south. This geometry suggests that the basement high was formed as an uplifted fault block with associated antithetic rotation of strata into the fault plane. c, Cross-section of the Woodlark basin just west of the propagating rift tip constructed from seismic profile 1218. d, Detail of the seismic profile showing Moresby seamount, an extremely large fault block¹ with a prominent fault plane forming its northern flank, dipping beneath the Woodlark rise. The association of this proposed fault plane with earthquake mechanisms that permit normal sense motion on low-angle planes in this region²⁸ supports the suggestion that Moresby seamount forms the footwall of a major low-angle normal fault²¹ along which the deep crust of the lower Woodlark rise has been exhumed. Moresby seamount and the large basement block imaged in our Goodenough basin seismic profile (b) are probably submerged core complex structures formed as a response to rifting. The Goodenough basin feature, like the core zones of the islands, must have formed earlier than Moresby seamount given the present relatively flat-lying sediment cover in Goodenough basin overlying the basement.

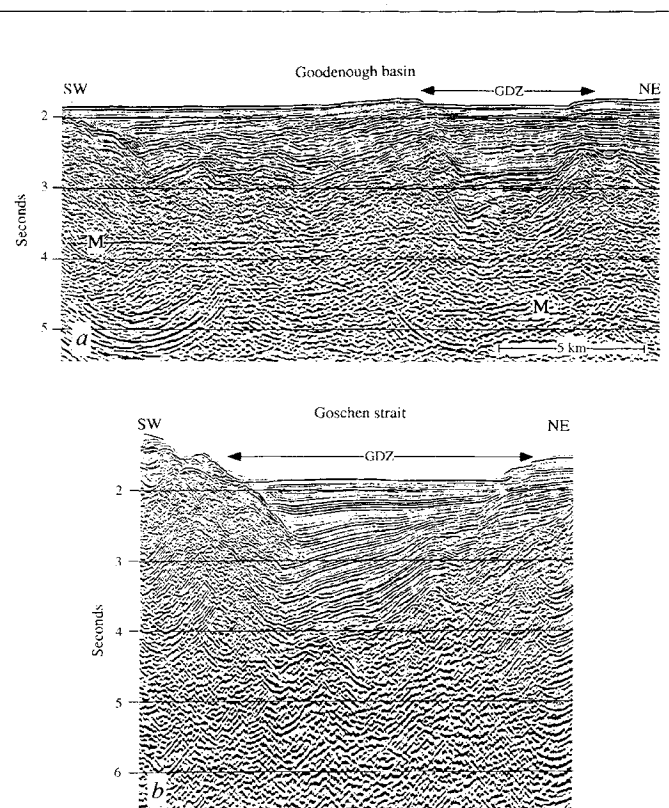


FIG. 3 a, b, Seismic profiles of the active Goschen deformation zone (GDZ), which we suggest to be an analogue of the dying rift of an oceanic propagator. a, Profile from the centre of the Goodenough basin (Fig. 1), where the GDZ forms a graben within a zone of fault blocks rotating along northeast-dipping fault surfaces (M is a multiple reflection). The GDZ widens to the east and merges with the half-graben of Goschen strait, shown in b. The strong southwest dip of sediments in Goschen strait has resulted from consistent localization of strain in this region, as compared to the more distributed deformation characteristic of the central and western Goodenough basin.

same amount of extensional strain as is occurring in the North and South valleys. In the MERZ, located between the east-west corridor of extensional deformation and Goschen strait (Fig. 4a), lithosphere that is instantaneously defined to reside north of the new plate boundary in GDZ is transferred to the Indo-Australian plate. The ridge and trough topography of the western Pocklington rise may reflect abandoned remnants of GDZ-like structures in much the same way as oceanic MERZs include fragments of failed ridges (compare Fig. 4a and b). The Woodlark

MERZ is aseismic at teleseismic detection levels, as are oceanic MERZs.

The oceanic-like behaviour of breakup propagation may result because the region undergoing extension was thickened by plate convergence in an arc-continent setting for around 40 Myr before extension began^{2,3} producing a thick, relatively hot, and hence weak lithosphere. Deformation of a thick, hot lithosphere provides the conditions necessary for development of metamorphic core complexes²⁹, and several such features are well

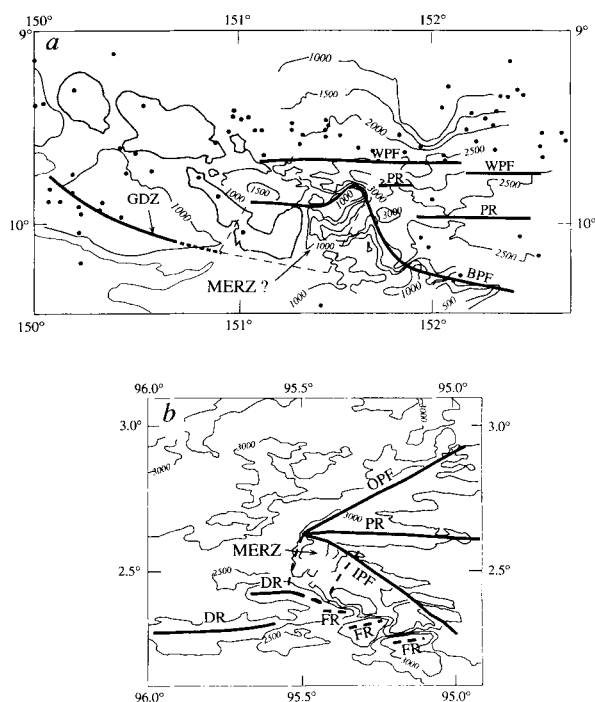


FIG. 4 Ocean-like behaviour of breakup. *a*, A summary diagram of major tectonic features of the Woodlark-D'Entrecasteaux region showing how breakup could be conceptualized in terms of an oceanic propagation system. *b*, Tectonic elements of the Galapagos propagator²⁶. In the propagating rift model, the Woodlark basin pseudo-fault (WBF) and Bonvouloir pseudo-fault (BPF in *a*) become analogous to the oceanic outer pseudo-fault (OPF) and inner pseudo-fault (IPF in *b*), respectively. A corridor of extreme, focused continental extension incorporating North and South valleys (Fig. 1) occurs west of the propagating rift (PR) tip. The Goodenough deformation zone (GDF) forms an opposing corridor of deformation comparable to a dying rift (DR in *b*). The dashed continuation of the GDZ into Goschen strait represents the probable continuation of deformation into this zone, as suggested from bathymetric considerations. Between these corridors lies a highly deformed province that could be thought of as a zone lithospheric transfer equivalent to an oceanic migrating extensional relay zone (MERZ) that rafts away as a very disturbed block incorporating limbs of the GDZ or DR, forming failed rifts (FR in *b*) in oceanic case.

described in the region. The strength profile of hot, thick continental lithosphere, in fact, resembles that of young oceanic lithosphere^{14,19,20,30,31}, and we conjecture that this similarity is what allows continental deformation in the region to mimic so closely that of an oceanic propagator. □

Received 6 July 1995; accepted 12 February 1996.

1. Taylor, B., Goodliffe, A., Martinez, F. & Hey, R. *Nature* **374**, 534–537 (1995).
2. Davies, H. L., Symonds, P. A. & Ripper, I. D. *Bur. Miner. Resour. J. Austr. Geol. Geophys.* **9**, 49–68 (1984).
3. Davies, H. L. *Am. J. Sci.* **280-A**, 171–191 (1981).
4. Davies, H. L. & Warren, R. G. *Tectonics* **7**, 1–21 (1988).
5. Hill, E. J., Baldwin, S. L. & Lister, G. S. *Geology* **20**, 907–910 (1992).
6. Hill, E. J. & Baldwin, S. L. *J. metamorph. Geol.* **11**, 261–277 (1993).
7. Lister, G. S. & Baldwin, S. L. *Geology* **21**, 607–610 (1993).
8. Baldwin, S. L., Lister, G. S., Hill, E. J., Foster, A. & McDougall, I. *Tectonics* **12**, 611–628 (1993).
9. Binns, R. A. et al. *Proc. Pacific Rim Congr.* **87**, 531–535 (1987).
10. Binns, R. A. et al. *Trans. Am. geophys. Un.* **67**, 1231 (1986).
11. Courtillot, V., Gaedeano, A. & LeMouél, J. L. *Earth planet. Sci. Lett.* **47**, 144–160 (1980).
12. Vick, G. E. *J. geophys. Res.* **87**, 10677–10688 (1982).
13. Courtillot, V. *Tectonics* **1**, 239–250 (1982).
14. Steckler, M. S. & tenBrink, U. *Earth planet. Sci. Lett.* **79**, 120–132 (1986).
15. Acton, G. D. & Stein, S. *Tectonics* **10**, 501–526 (1991).
16. Courtillot, V. et al. *J. geophys. Res.* **89**, 3315–3333 (1984).
17. Weissel, J. K., Taylor, B. & Karner, G. D. *Tectonophysics* **87**, 253–277 (1982).
18. Taylor, B. in *Marine Geology, Geophysics and Geochemistry of the Woodlark Basin—Solomon Islands* (eds Taylor, B. & Exon, N. F.) 25–48 (Circum-Pacific Council for Energy & Mineral Resources, Houston, Texas, 1987).
19. Buck, W. R. *J. geophys. Res.* **96**, 201610–20178 (1991).
20. Froidevaux, C. & Isacks, B. *Earth planet. Sci. Lett.* **71**, 305–314 (1984).
21. Mutter, J. C. et al. *Trans. Am. Geophys. Un.* **73**, 536 (1992).
22. Goodliffe, A., Taylor, B., Martinez, F. & Hey, R. *Trans. Am. geophys. Un.* **74**, 606 (1994).
23. Bennes, V., Scott, S. D. & Binns, R. A. *J. geophys. Res.* **96**, 4439–4456 (1994).

24. Mutter, J. C., Mutter, C. Z., Abers, G. A. & Fang, J. *Trans. Am. geophys. Un.* **74**, 412 (1993).
25. Hey, R. N. *Earth planet. Sci. Lett.* **37**, 321–325 (1977).
26. Hey, R. N., Kleinrock, M. C., Miller, S. P., Atwater, T. M. & Searle, R. C. *J. geophys. Res.* **91**, 3369–3393 (1986).
27. Searle, R. & Francheteau, J. *Mar. Geophys. Res.* **8**, 95–129 (1986).
28. Abers, G. A. *Geology* **19**, 1205–1208 (1991).
29. Lister, G. S., Etheredge, M. A. & Symonds, P. A. *Geology* **12**, 246–250 (1986).
30. England, P. C. *Geophys. J. R. astr. Soc.* **70**, 295–321 (1982).
31. Mutter, J. C. *Nature* **374**, 499–500 (1995).

ACKNOWLEDGEMENTS. We thank the crew and technical staff of the R/V *Maurice Ewing* for their considerable efforts allowing the acquisition of multichannel seismic data in reef-laden waters. We thank Western Geophysical, Inc. for providing the Omega seismic processing software used to process the seismic images, and S. M. Carbotte, M. K. McNutt and B. Taylor for reviews. This work was supported by the US NSF.

Bitumen as a hafting material on Middle Palaeolithic artefacts

Eric Boëda*, Jacques Connan†, Daniel Dessort†, Sultan Muhesen‡, Norbert Mercier§, Hélène Valladas§ & Nadine Tisnérat§

* Université de Paris X, Nanterre, Département d'Ethnologie/Préhistoire, 200 Avenue de la République, 92001 Nanterre Cedex, France

† Elf Aquitaine, Direction Exploration, CSTJF, Avenue Larribau, 64018 Pau Cedex, France

‡ Ministère de la Culture, Direction Générale des Antiquités et des Musées, Damas, Syria

§ Centre des Faibles Radioactivités, Laboratoire mixte CNRS-CEA, 91198 Gif-sur-Yvette Cedex, France

A SCRAPER and a Levallois flake, discovered in the Mousterian levels (dated around 40,000 BC) of the Umm el Tiel site in Syria, were submitted to an organic geochemical study to identify a black substance occurring on their surface. The shape of this black substance suggests that the organic traces are remnants of a hafting material used by Middle Palaeolithic people to glue handles onto their tools. Gas chromatography–mass spectrometry (GC–MS) analyses of both C_{15} alkanes and C_{15+} aromatics confirm that the black substance is a highly weathered bitumen, the source of which remains unknown. According to some diagnostic molecular information (for example, the occurrence of fluoranthene and pyrene), it seems that the raw bitumen used has been subjected to extreme temperature. The scraper and the Levallois flake described here are, to the best of our knowledge, the first reported examples of Middle Palaeolithic artefacts hafted with bitumen to handles.

Middle Palaeolithic flint tools are usually considered to be rudimentarily made using unsophisticated techniques. It would

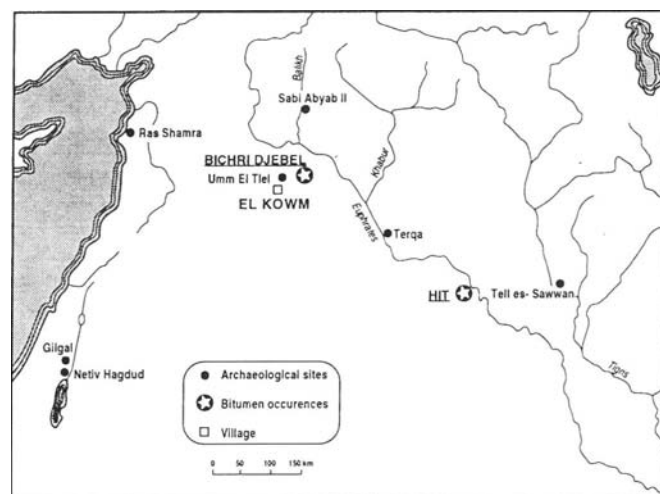


FIG. 1 Location of archaeological sites and bitumen sources.

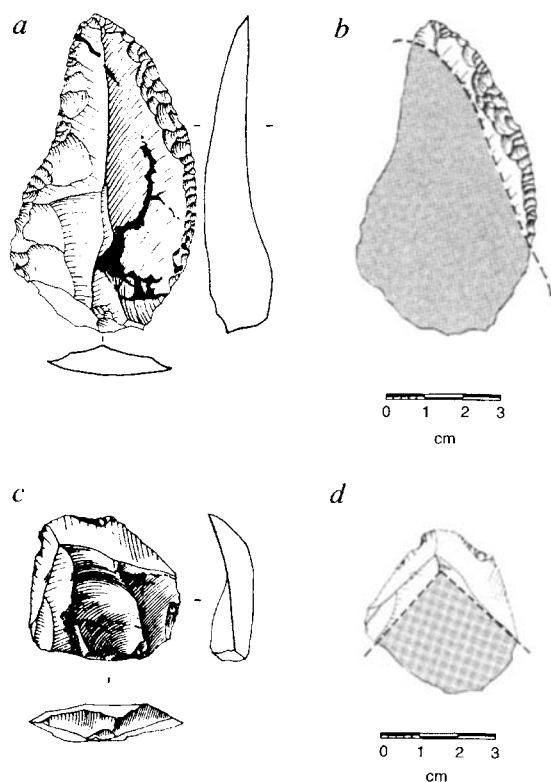


FIG. 2 *a*, Drawing showing the black bitumen traces on the upper face of the convergent scraper. On this face, the sinuous trace goes across the scraper and follows the convex curve of the right edge, at a distance of 1 cm. In the proximal third, the trace swerves to end on the right edge. *b*, Reconstruction of the presumed handle of the convergent scraper. *c*, Drawing showing the black bitumen traces on the upper face of the Levallois flake, made on a subquadrangular, 4-cm long and 4-cm wide, small piece. The black traces are only on the face shown. There are small marks, creating three parallel lines to the edges at a distance of 1 cm. *d*, Reconstitution of the presumed handle, only valid for the limits of the handle on the artefact.

be surprising if Middle Palaeolithic people used more complex procedures such as hafting. Studies based on micro-trace analyses^{1,2}, however, have shown that handles did exist at those times, although there has been very little evidence of them. As far as we know, until now there has been no evidence that glue was used in the manufacture of such tools.

We have discovered traces of bitumen on artefacts from the most recent Mousterian layers of the open-air site of Umm el Tiel in the El Kowm basin, located in the Syrian desert between Palmyra and the Euphrates river (Fig. 1). The outlines of these traces suggest different types of hafting, and physicochemical analysis indicates that the bitumen was heated before use.

The first use of bitumen³⁻⁶ or other adhesive materials^{7,8} had previously been thought to date from the Neolithic, as at Netiv Hagdud near the Dead Sea (8,200–7,500 bc) and at Tell Sabi Abyad II⁹ (PPNB, 6,300 bc). The discoveries at Umm el Tiel show that bitumen was used as an adhesive more than 40,000 years ago. These new data suggest that Palaeolithic people had greater technical ability than previously thought, as they were able to use different materials to produce tools.

The site of Umm el Tiel was discovered in 1978 (refs 10 and 11), and its Neolithic levels were first excavated in 1989 (ref. 12). The Palaeolithic levels were excavated in 1991 (ref. 13) and have already yielded 51 layers, with Aurignacian, Transition and Mousterian industries. The penultimate layer in the Mousterian sequence (layer IV2a) has yielded a convergent side-scraper on a

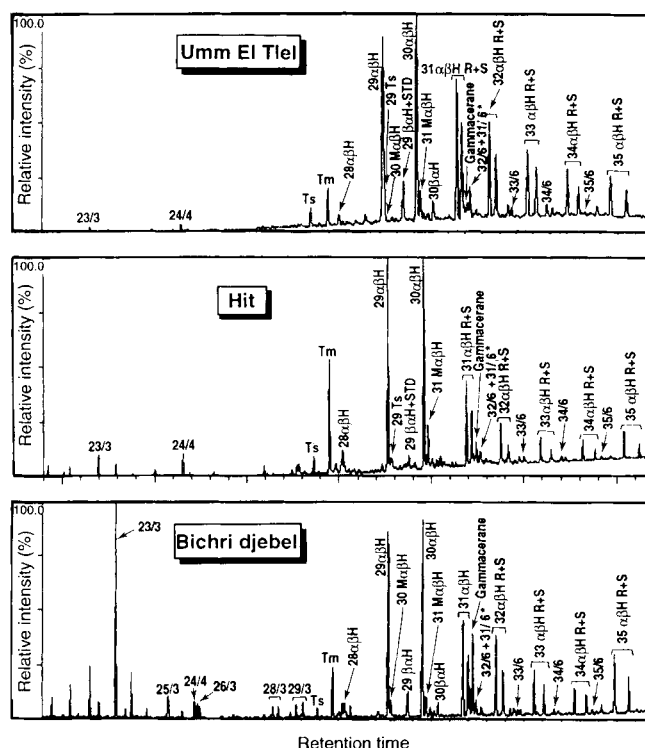


FIG. 3 Terpene distribution pattern (m/z 191) of the chloroform extract of the Umm el Tiel scraper, compared to characteristic fingerprints of two well-known natural asphalts: Abu Gir in the Bichri djebel in Syria, and Hit along the Euphrates river in Iraq. Terpenes from the Bichri djebel oil-stained sands contain significant tricyclopolyrenanes but much higher Tm-to-Ts and gammacerane-to-hopane ratios. Abbreviations: Tm, C27 α -hopane; Ts, C27 α -neohopane; 29 α H, C29 α -hopane; 30M α H, 2 α -methyl-C29 α -hopane; 23/3, C23 tricyclopolyrenane; 24/4, C24 tetracyclic terpene; 33/6, C33 hexahydrobenzohopane; 31/6*, the novel C31hexahydrobenzohopane, referred to as structure **6** in ref. 24.

Levallois flake and a quadrangular Levallois flake, both of which have traces of bitumen (Fig. 2).

The trace of bitumen is present on both faces of the side-scraper and, except at the point, it follows the curve of the right side of the tool, 1 cm below its edge (Fig. 2*a*). The left side and the proximal end of the tool were set into a handle. This functional association of a side and a point seems to be a constant trend in Mousterian assemblages (Fig. 2*b*).

Bitumen can be seen on only one face of the quadrangular Levallois flake, mainly near the middle (Fig. 2*c, d*). We suggest that the handle was fastened to the middle part of the tool, leaving both sides free for cutting and scraping.

Precise dating of layer IV2a has not been possible, but thermoluminescence and ¹⁴C dates are available for the overlying layers, which belong to transition industries (layers III) and the Aurignacian (layers II). The thermoluminescence of three burnt flints¹⁴ from layer III2a gave a mean age of 36,000 $\pm$ 2,500 years; an accelerator mass spectrometry (AMS) radiocarbon date¹⁵ of 34,530 $\pm$ 890 yr before present (BP) (laboratory reference, GifA 93216) has been obtained for the same layer. The thermoluminescence of four burnt flints from the overlying layer II date it at 34,000 $\pm$ 2,500 yr BP; an AMS radiocarbon date of 32,000 $\pm$ 580 yr BP (GifA 93212) has also been obtained for the same layer.

Sonic extraction with hot chloroform was used to determine the type of glue on the two Umm el Tiel artefacts. The resulting extract was then fractionated to isolate C₁₅₊ alkanes and C₁₅₊ aromatics. Finally, both fractions underwent gas chromatography and computerized GC-MS analysis to determine their molecular spectra.

Sterane and terpane distribution patterns, analysed by GC-MS, confirm the fossil origin of the black substance (Fig. 3). Terpanes are dominated by $\alpha\beta$ -hopanes and 2α -methyl- $\alpha\beta$ -hopanes¹⁶, with hexahydrobenzohopanes as subordinate families. Note the very low quantities of tricycloprenanes¹⁷ and gammacerane coupled with a Tm-to-Ts ratio of about 1.5. (Tricyclopolyrenanes, gammacerane, Tm and Ts are biomarkers or geochemical fossils from the alkanes which are used as genetic parameters to trace the origin of bitumens.)

Most of the steranes are dominated by regular steranes, but diasteranes (or rearranged steranes) and methylsteranes are also present. Unfortunately, the distributions of steranes and terpanes in the Umm el Tiel artefacts do not display the characteristic fingerprints of the most well-known bitumen occurrences in these areas, namely the Abu Gir tar sands in the Bihri Deibel in Syria, and the Hit deposit in Iraq (Figs 1 and 3).

The GC-MS analysis of C₁₅₊ aromatics from Umm el Tiel has allowed us to identify several classes of diagnostic compounds. (1) Linear T-alkylbenzenes were first detected on a sickle (1,500 BC) from Ras Shamra¹⁸ (Fig. 1) in Syria, and were fully identified in a bitumen-impregnated mat (5,500–5,000 BC)

from Tell es-Sawwan¹⁹ (Fig. 1) in Iraq. These molecular families, initially thought to be potential contaminants introduced by detergents²⁰, are widespread in bituminous remains from various archaeological sites. In these bitumen mixtures, there is generally evidence of pyrolytic products. The associated molecular characteristics suggest that the T-alkylbenzenes are true indigenous products, which may have been formed by Fischer-Tropsch reactions of a long-chain alkene with benzene during the thermal treatment of bitumens. This hypothesis is substantiated by heating experiments on coal from the Mahakam delta, in which T-alkylbenzenes are known to be the dominant alkylbenzenes²¹, generated at temperatures exceeding 350 °C. (2) Fluoranthene, pyrene, phenanthrenes and chrysenes are characteristic pyrolytic products²². (3) Benzohopanes, monoaromatic steroids and 8, 14-secohopanoids are typical biomarkers, currently identified in natural asphalts and archaeological bitumens from the Near East²³.

In conclusion, these two artefacts exhibit traces of bitumen that was heated before being used as a glue to secure the hafting of a handle. As far as we know this is the earliest example of bitumen being treated and used by man. □

Received 31 July 1995; accepted 24 January 1996.

- Anderson-Gerfaud, P. thesis, Univ. Bordeaux (1981).
- Beyries, S. in *La main et l'outil: manches et emmanchements préhistoriques* 55–61 (Travaux de la Maison de l'Orient, No 15, G. S. Maison de l'Orient, 1987).
- Barquins, M. *Science* **192**, 8–10 (1993).
- Connan, J. & Deschesne, O. in *Materials Issues in Art and Archaeology* Vol. III (eds Vandiver, P. B., Druzik, J.-R., Wheeler, G.-S. & Freestone, I.-C.) 683–720 (Material Research Society, Pittsburgh, PA, 1992).
- Connan, J. & Deschesne, O. *Recherche* **229**, 152–159 (1991).
- Lechevallier, M. in *Les cités oubliées de l'Indus-Archéologie du Pakistan. Catalogue de l'Exposition au Musée Guimet* (ed. Jarrige, J.-F.) 56–57 (Association Française d'Action Artistique, Paris, 1988).
- Binder, D., Bourgeois, G., Benoist, F. & Vitry, C. C. *Rev. Archéomét.* **14**, 37–42 (1990).
- Funk, H. thesis, Hamburg Univ. (1969).
- Verhoeven, M. *Orient Express* 9–12 (1994).
- Fujimoto, T. in *Paleolithic Site of Douara Cave and Paleogeography of Palmyra Basin in Syria* (eds Hanihara, K. & Akazawa, T.) 77–130 (Univ. Tokyo Press, 1979).
- Cauvin, J., Cauvin, M.-C. & Stordeur, D. *Cah. Euphrate* **2**, 80–118 (1979).

- Molist, M., Cauvin, M.-C. & Taha, A. *Ann. archéol. arab. syrien.* **37/38**, 67–77 (1990).
- Boëda, E. & Muhesen, S. *Cah. Euphrate* **7**, 47–60 (1993).
- Mercier, N., Valladas, H. & Valladas, G. *Quat. Sci. Rev.* **14**, 351–364 (1995).
- Arnold, M., Bard, E., Maurice, P. & Duplessy, J.-C. *Nucl. Instrum. Meth. Phys. Res.* **29**, 120–123 (1987).
- Summons, R.-I. & Jahnke, L.-L. *Geochim. cosmochim. Acta* **54**, 247–251 (1990).
- Aquito Neto, F.-R., Trendel, J.-M., Restlé, A., Connan, J. & Albrecht, P. in *Advances in Organic Geochemistry* (eds Bjoroy, M. et al.) 659–667 (Wiley, Chichester, 1993).
- Connan, J., Deschesne, O. & Dessort, D. *Bull. Centres Rec. Explor. Prod. Elf-Aquit.* **14**, 2, 521–539 (1990).
- Connan, J., Dessort, D. & Deschesne, O. *Bull. Centres Rec. Explor. Prod. Elf-Aquit.* **16**, 1, 33–53 (1992).
- Eganhouse, R.-P. *Int. J. envir. analyt. Chem.* **26**, 241–263 (1986).
- Mansuy, L. thesis, Inst. Poly. Lorraine, Nancy (1995).
- White, R. *Nat. Gallery techn. Bull.* **10**, 58–71 (1996).
- Connan, J. & Deschesne, O. *Mesopotamia* **XXVII**, 47–64 (1992).
- Schaeffer, P., Adam, P., Trendel, J.-M., Albrecht, P. & Connan, J. *J. Org. Geochem.* **23**, 87–89 (1995).

Fitness costs of resource overlap among coexisting bird species

Thomas E. Martin

U.S. National Biological Service, Montana Cooperative Wildlife Research Unit, University of Montana, Missoula, Montana 59812, USA

NATURAL selection is thought to favour differences in resource use (resource partitioning) among coexisting species because overlap is assumed to incur fitness costs^{1–5}. However, studies of resource partitioning in animal systems in the wild have ignored or had difficulty demonstrating fitness costs of overlap, particularly at the level of individuals. Here I report manipulative and observational tests demonstrating individual-level fitness costs of resource overlap. Based on an extensive data set (2,400 experimental nests and 1,408 natural nests), I show that seven coexisting bird species differ in nesting microhabitats, and that overlap in use of nest sites increases nest predation rates (fitness costs). Moreover, predation risk is greater for individuals within species that use nest sites that overlap with coexisting species. Such intraspecific variation can allow natural selection to shape species differences and patterns of species coexistence.

Competition for food is often assumed to be the cause of

resource partitioning, but the risk of nest predation can theoretically increase with overlap in resource (nest site) selection, and thereby favour resource partitioning among coexisting species^{6–12}. The risk of nest predation may increase with the overlap of nest sites because the resulting increase in cumulative nest density in a microhabitat can reward predators which search similar microhabitats^{10–12}. Nest predation is the primary source of nesting mortality for birds, and so it can exert strong selection on traits such as the choice of nest site^{10–16}. Nest sites represent microhabitat choice by individuals, and so allowed me to examine the fitness (predation) costs of individual differences in resource overlap among coexisting species. I used both artificial and real nests to test whether predation increases with overlap among species.

Artificial nests baited with real eggs were placed in sites typical of bird species that occur on high-elevation forest sites in Arizona¹⁷. Results of experiments over a period of six years showed that predation rates were higher when species coexisted than when they were separate (Fig. 1). In addition, predation rates increased more when species were similar in nest placement than when they differed (Fig. 1). These results using artificial nests suggest that predation can select against the coexistence of species that are similar in their selection of nest sites.

To examine nest-site partitioning, I studied the microhabitat of nest sites of all ground- and shrub-nesting bird species on the Arizona field sites for eight years. Species differed markedly in the habitat characteristics of their nest sites for both ground- and shrub-nesting species (Fig. 2). Individuals within species, however,

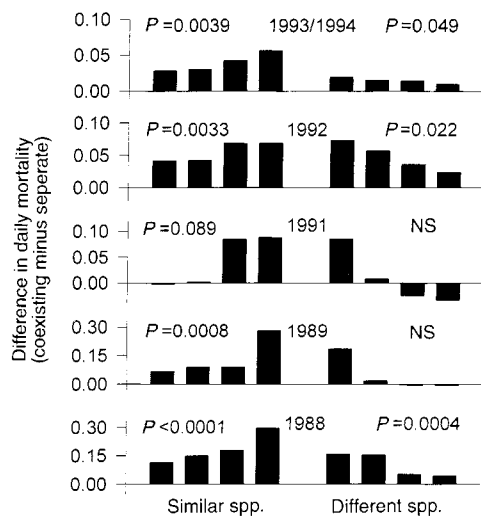


FIG. 1 Differences in daily mortality rates (probability of predation per day)²³ for artificial nests baited with quail (*Coturnix coturnix*) eggs for nests placed in sites simulating two species together (coexisting) and each species separately. Differences in daily mortality rates (DMRs) were calculated by

subtracting DMR when coexisting from DMR when separate. Nests were placed in sites that simulated those of actual bird species. Nests were placed in the design: X X X O X O X O O O, where X and O represent different species, and each replicate included 40, 80 and 40 nests (that is, 160 nests per replicate per year), respectively, in 1988 and 1989, and 20, 40 and 20 nests (that is, 80 nests per replicate per year), respectively, in remaining years. Conspecific nest density, as measured by the distance between nests, was held constant (40m in 1988 and 1989, and 80 m for remaining years) for species when alone and coexisting, but total nest density was doubled (that is, distance between heterospecifics was half that of conspecifics) when the two species coexisted (X O X O X O) compared to when separate (X X X or O O O). These experiments were replicated across four sites (bars) for six years for each of two treatments (similar versus different species) and included a total of 2,400 artificial nests. For the similar species, the two simulated nest types (X and O) were similar (X and O were both ground-nesters or were both shrub-nesters), and for different species, the two simulated nest types were different (X was a ground-nester and O was a shrub- or subcanopy-nester). Differences in daily mortality rates when coexisting versus separate were tested using the program CONTRAST²⁴; P-values show whether nest predation rates were greater when coexisting than when separate for each treatment in each year. Increases in nest predation were greater when species were similar (mean DMR increase $\pm$ s.e.m. = 0.091 ± 0.018) than when species were different (D.M.R. increase = 0.044 ± 0.014) (analysis of variance (ANOVA): $F = 6.2$, $P = 0.018$ for treatments; $F = 5.7$, $P = 0.002$ for year; $F = 0.7$, $P = 0.60$ for treatment by year).

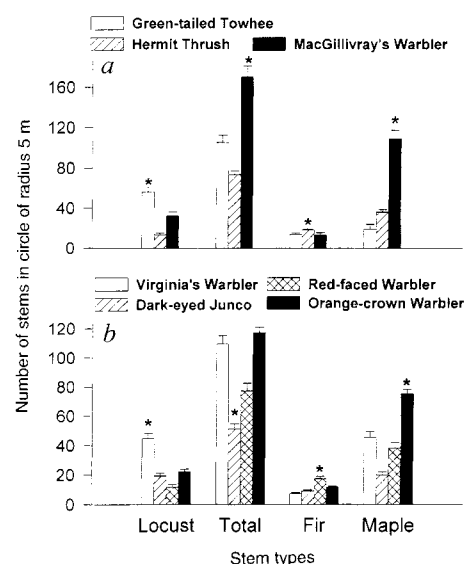
varied in the extent to which they overlapped with the nest sites of other coexisting species. This variation among individuals allowed me to examine whether nest predation increased with microhabitat overlap in a natural system. The results showed that nest predation was indeed higher for individuals with greater overlap of nest sites for six of the seven species (Fig. 3). Thus a strong fitness cost is incurred by being similar in nest-site selection, and such costs can favour resource partitioning.

The most commonly used example of resource partitioning in textbooks of biology¹⁸, ecology¹⁹ and evolution²⁰ is a study by MacArthur²¹, which showed that coexisting birds (warblers) differed in foraging microhabitats. This study²¹ is used as a classic example of the effects of competition favouring resource partitioning, but the fitness costs of overlap (and, hence, competition) were never tested¹⁰. Although this example is widely used, it is not

unusual; fitness costs of resource overlap are rarely tested in the wild, particularly for individuals.

Here I show, by using tests with both artificial (also ref. 12) and natural nests, that individuals using microhabitats that overlap with coexisting species incur fitness costs from nest predation. The increase in nest predation for individuals that overlap with other species could be due to these individuals using nest sites that are outside their habitat optima, rather than being due to overlap *per se*. However, the end result is the same, because individuals that use nest sites that overlap with those of other species incur fitness costs. This intraspecific variation is particularly significant if selection of nest sites is heritable, because then such variation provides the source for natural selection to favour observed differences in nest-site selection among coexisting species. Microhabitat differences, however, could be set by alternative influences

FIG. 2 Differences in microhabitat (numbers of stems in a circle of radius 5 m) among bird species within a, shrub-nesting, and b, ground-nesting guilds. Nests were located by following parents to reduce biases of finding obvious nests that can result from random search²⁵. Four microhabitat variables (numbers of stems of canyon maple *Acer grandidentatum*, New Mexican locust *Robinia neomexicana*, white fir *Abies concolor*, and all woody plants) account for most of the overall variation in habitat of the study sites¹⁰. These four variables, together with the plant species under which or in which nests were placed, were used in discriminant function analysis to examine whether species differed in nest-site characteristics¹⁶. The resulting discriminant functions were highly significant (Wilks' Lambda, $P < 0.0001$), and species differed strongly in nest-site characteristics within the canonical space defined by these discriminant functions, as tested by Mahalanobis distances ($P < 0.0001$ for all pair-wise comparisons of species within shrub- and ground-nesting guilds). Among the shrub-nesters, green-tailed towhees (*Pipilo chlorurus*) had more New Mexican locust, MacGillivray's warblers (*Oporornis tolmiei*) had more canyon maple and total stems, and hermit thrushes (*Catharus guttatus*) had more small firs around their nests than the other species. Among the ground-nesters, orange-crowned warblers (*Vermivora celata*) had more canyon maple, Virginia's warbler (*Vermivora virginiae*) had more New Mexican locust, and red-faced warblers (*Cardellina rubrifrons*) had more small white firs around their nests than the other bird species. Dark-eyed juncos (*Junco hyemalis*) placed most nests in the open under grass clumps and had fewer total woody stems than the other species. These species-specific differences are highlighted by asterisks.



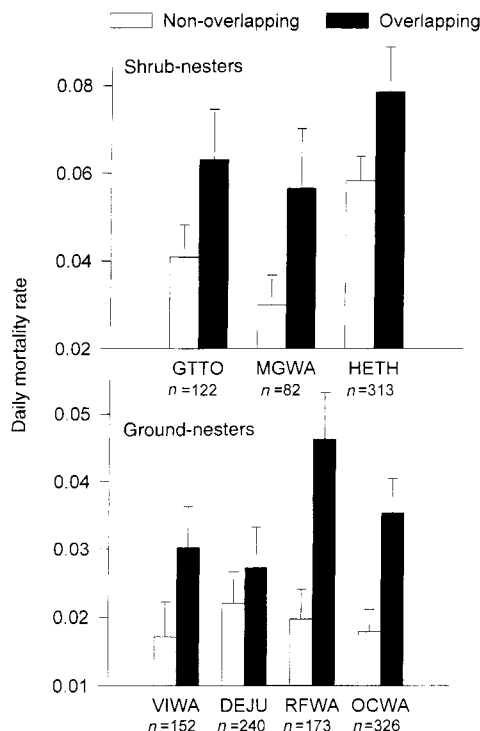


FIG. 3 Daily mortality rates (predation probability per day)²³ for nests of individuals that overlapped or did not overlap coexisting species based on nest-site characteristics (see Fig. 2). Nest-site characteristics were used to classify identity of the bird species predicted to occupy the nest site based on discriminant function analysis of nest microhabitat characteristics (see Fig. 2). By definition, nest sites in which the owner bird species was incorrectly classified are nest sites that overlapped with other species in habitat characteristics. Daily mortality rates were calculated for nest sites that were incorrectly predicted (overlapping sites) and compared to nest sites that were correctly predicted to owner species (non-overlapping sites). Differences in daily mortality rates between overlapping and non-overlapping sites were tested using the program CONTRAST²⁴. This program is appropriate for testing homogeneity of maximum-likelihood estimates, such as daily mortality rates, and uses a chi-squared approach that is analogous to ANOVA to control for experimentwise error and adjust for type I errors^{24,26}. Daily mortality rates were greater ($P < 0.001$ in all cases) for nests that overlapped with other species in nest-site characteristics than for non-overlapping nest sites in all cases except dark-eyed juncos. GTTO, green-tailed towhee; MGWA, MacGillivray's warbler; HETH, hermit thrush; VIWA, Virginia's warbler; DEJU, dark-eyed junco; RFWA, red-faced warbler; OCWA, orange-crowned warbler.

over the evolutionary histories of species^{11,22}. Regardless of the cause of microhabitat differences, overlap of nest sites incurs predation costs that can influence the reproductive success of species in differing assemblages and so ultimately influence the patterns of coexistence. □

Received 12 October 1995; accepted 17 January 1996.

- Lack, D. *Ecological Isolation in Birds* (Harvard Univ. Press, Cambridge, MA, 1971).
- Schoener, T. W. *Science* **185**, 27–39 (1974).
- Schoener, T. W. in *Community Ecology* (eds Kikkawa, J. & Anderson, D. J.) 91–126 (Blackwell Scientific, Melbourne, 1986).
- Gause, G. F. *The Struggle for Existence* (Williams & Wilkins, Baltimore, 1934).
- Hutchinson, G. E. 1957 *Cold Spring Harb. Symp. quant. Biol.* **22**, 415–427.
- Ricklefs, R. E. & O'Rourke, K. *Evolution* **29**, 313–324 (1975).
- Holt, R. D. *Theor. Populat. Biol.* **12**, 197–229 (1977).
- Holt, R. D. *Am. Nat.* **124**, 377–406 (1984).
- Jeffries, J. J. & Lawton, J. H. *Biol. J. Linn. Soc. Lond.* **23**, 269–286 (1984).
- Martin, T. E. *BioScience* **43**, 523–532 (1993).
- Martin, T. E. *Evol. Ecol.* **2**, 37–50 (1988).
- Martin, T. E. *Proc. natn. Acad. Sci. U.S.A.* **85**, 2196–2199 (1988).
- Ricklefs, R. E. *Smithson. Contr. Zool.* **9**, 1–48 (1969).
- Martin, T. E. *Am. Nat.* **141**, 897–913 (1993).
- Martin, T. E. *Ecol. Monogr.* **65**, 101–127 (1995).
- Martin, T. E. & Roper, J. J. *Condor* **90**, 51–57 (1988).
- Martin, T. E. *Ecology* **68**, 74–84 (1988).

- Solomon, E. P., Berg, L. R., Martin, D. W. & Vilee, C. *Biology* (Saunders College Publishing, New York, 1993).
- Brewer, R. *The Science of Ecology* (Saunders College Publishing, New York, 1994).
- Strickberger, M. W. *Evolution* (Jones & Bartlett, Boston, MA, 1996).
- MacArthur, R. H. *Ecology* **39**, 599–619 (1958).
- James, F. C., Johnston, R. F., Wamer, N. O., Niemi, G. J. & Boecklen, W. J. *Am. Nat.* **124**, 17–47 (1984).
- Hensler, G. L. & Nichols, J. D. *Wilson Bull.* **93**, 42–53 (1981).
- Hines, J. E. & Sauer, J. R. *U.S. Fish Wildl. Serv. Fish. Wildl. Tech. Rep. Vol. 24* (1989).
- Martin, T. E. & Geupel, G. R. *J. Fid Orn.* **64**, 507–519 (1993).
- Sauer, J. R. & Williams, B. K. *J. Wildl. Mgmt.* **53**, 137–142 (1989).

ACKNOWLEDGEMENTS. I thank many field assistants for help in collecting the field data; A. Badyaev, I. J. Ball, D. Barber, S. Garner, C. Ghalambor, P. Martin and J. Tewksbury for discussion and critiques; the Arizona Game and Fish Agency, Blue Ridge Ranger Station of the Coconino National Forest, and the Apache-Sitgreaves National Forest for their support of this work. This work was supported by grants from the National Science Foundation and the BIRD (Breeding Biology Research and Monitoring Database) program under the Global Change Research Program of the National Biological Service.

A hydrodynamic model of locomotion in the Basilisk Lizard

J. W. Glasheen*† & T. A. McMahon‡

† Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts 02138, USA

‡ Division of Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, USA

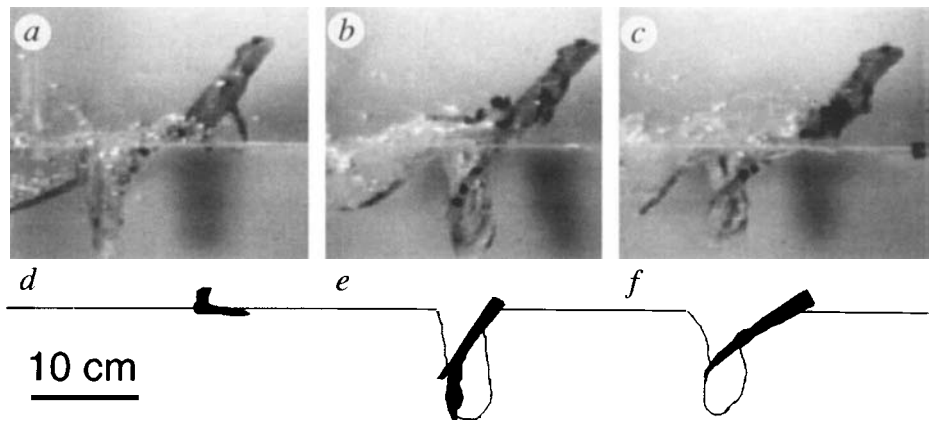
ORGANISMS with a body mass of more than one gram and which live at the air–water interface generally support their weight with their buoyant bodies. The maximum swimming speed these animals can attain is limited by wave-making resistance^{1–3}. For high-speed progression across a body of water, shore birds and basilisk lizards (*Basiliscus basiliscus*) support their bodies above the water surface by repeatedly striking the surface with their feet. Here we investigate the mechanism of support in moderately sized basilisk lizards (about 90 g) by combining hydrodynamic measurements of a physical model of the lizards' feet with an analysis of video records of foot movements. We find basilisks of intermediate size obtain little support for their body weight by slapping the water surface; most of the support comes from stroking the foot downwards while expanding an air cavity underwater. The lizard minimizes downward forces by pulling its foot upward before the cavity collapses.

Basilisk lizards (*Basiliscus basiliscus*) are well known for their ability to run on water^{4–6}. Yet there is no fundamental mechanistic understanding of how they are able to support their weight by running. We present a hydrodynamic model of weight support which consists of three phases: (1) slap, (2) stroke, and (3) protraction. A step begins as a lizard swings its foot through the air and slaps the water surface (Fig. 1a,d). Next, as the lizard strokes downwards, an air cavity is produced when air rushes in behind the foot (Fig. 1b,c). Finally, at the end of the stroke, the lizard pulls its foot upwards (protraction) within the air cavity and prepares for the next step (Fig. 1c,f).

To determine the hydrodynamic forces produced as the lizard runs across the water surface, we made detailed measurements of the lizards' movements during water-running sequences and conducted hydrodynamic experiments with a physical model of the lizards' feet. The hydrodynamic experiments allowed us to characterize three parameters: (1) m_{virtual} , (2) C_D^* , and (3) T_{seal} (refs 7, 11). The virtual mass associated with the foot at the moment of impact (m_{virtual}) was measured by dropping the foot model into the water at various impact speeds (Fig. 2a). The water-entry drag coefficient (C_D^*) of the foot was calculated by measuring the forces produced as the foot model descends and expands an air cavity underwater (Fig. 2b). Finally, the time between impact and

* Present address: 5130 Valley Life Sciences Building, University of California, Berkeley, Berkeley, California 94720, USA.

FIG. 1 A sample step of an 87.7-g lizard. *a*, The lizard's foot is parallel to the water surface at the moment of water impact (0 ms). An upward impulse (slap impulse) is produced as the lizard slaps the water surface and suddenly accelerates a volume of water downwards. *b*, After impact, an air cavity is produced as the lizard strokes downwards and backwards while plantar-flexing its foot (44 ms). Integrated over time, the upward component of the drag force helps to support the lizard's body weight (stroke impulse). *c*, Finally, the lizard minimizes the downward forces associated with foot protraction by pulling its foot upwards before the cavity collapses (68 ms). Protraction drag is further minimized by the feathering of the lateral toe fringes and the reorientation of the foot such that its long axis is parallel with the direction of movement. *d–f* Tracings of the lizard's left foot from the images in *a–c*, respectively. Seven basilisk lizards (*Basiliscus basiliscus*; 89.3 ± 6.1 g) were captured in Golfito, Costa Rica and brought to Cambridge, MA, USA. In a series of experiments, lizards were placed in a warm enclosure (30–35 °C) at one end of a long glass tank (3.66 m long by 0.43 m wide by 0.30 m high). A lizard was stimulated to run across the water surface by opening one side of



the enclosure while startling the lizard from the opposite side. By analysing high-speed video recordings of a lizard's movements (Kodak HSV, 500 fields per s), a plot of foot displacement with respect to time was obtained. Seventh-order polynomial fits of the displacement plots were analysed to determine foot angle ($\phi(t)$), mid-foot velocity ($u(t)$), and mid-foot depth ($h(t)$) as a function of time (LabView 2.2.1, National Instruments).

cavity closure (T_{seal}) was determined over a range of speeds and compared with measurements of the time between the impacts of the lizards' steps (Fig. 2c).

With the hydrodynamic parameters of the foot characterized, the forces a foot produces can be calculated. The force impulse produced when a lizard's foot slaps the water surface (slap impulse) is the product of the virtual mass associated with the foot (m_{virtual}) and the velocity to which the virtual mass is accelerated (u_{peak}),

$$\text{Slap impulse} = m_{\text{virtual}} u_{\text{peak}} \quad (1)$$

As the foot strokes downwards, both hydrostatic pressure and

a pressure due to the inertia of water create a drag force (Drag(t)). The component of the drag force in the vertical direction is obtained by multiplying Drag(t) by the cosine of the angle between a normal to the foot and the vertical ($\phi(t)$). Finally, the stroke impulse is the integral of the upward drag force over time,

$$\text{Stroke impulse} = \int \text{Drag}(t) \cos \phi(t) dt \quad (2)$$

We assume that as long as a lizard lifts its foot before the air cavity collapses, downward forces produced during foot protraction can

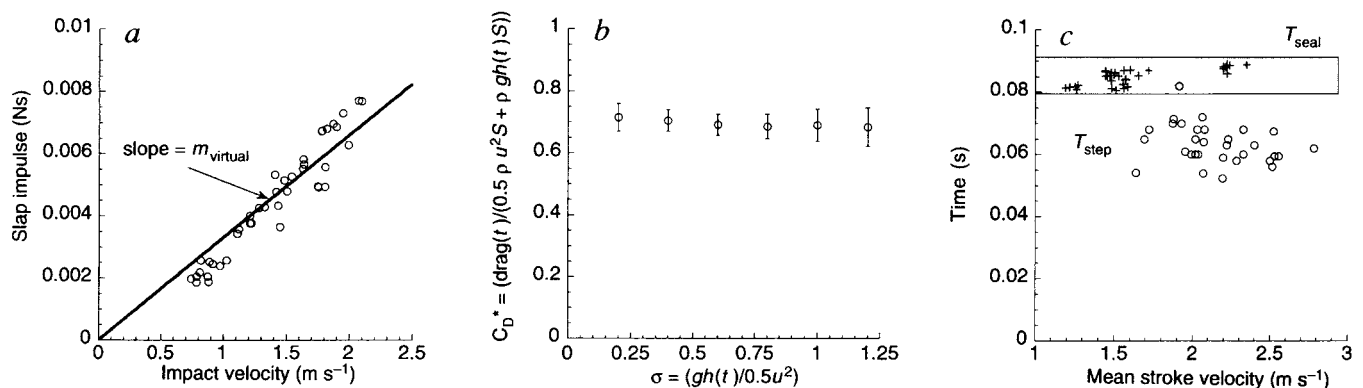
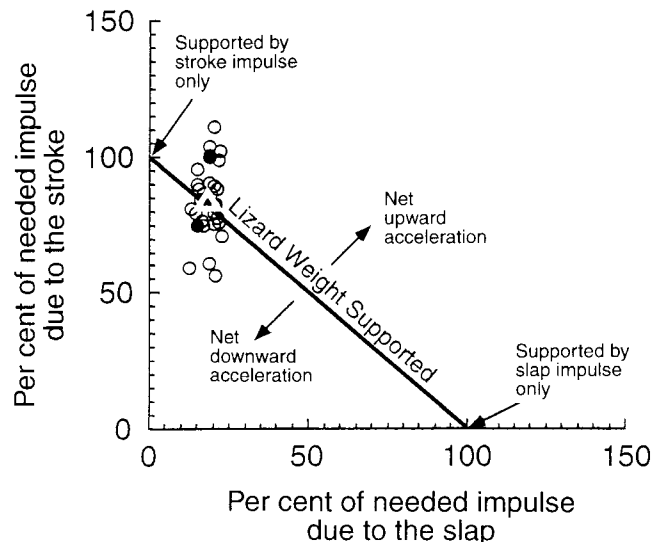


FIG. 2 Measurements obtained from two sets of hydrodynamic experiments with a model foot. *a*, The first set of experiments measured the slap impulse generated when the model foot strikes a water surface. The impulse increased linearly with rising impact speed. The slope of the line relating slap impulse to impact speed is a measure of the virtual mass (m_{virtual}) associated with the foot impacting the water surface ($m_{\text{virtual}} = 3.30$ g). *b*, The second set of experiments measured the drag force generated as the foot model expands an air cavity underwater. As has been found for disks¹¹, the dimensionless force (C_D^* ; mean $\pm$ s.d.; 39 measurements; overall mean = 0.680 ± 0.0310) that is produced as the model foot strokes downward at nearly constant velocity is invariant over a broad range in dimensionless depth ($gh(t)/(0.5u^2)$; g is gravitational acceleration, $h(t)$ is the depth of the model foot below the water surface, and u is the velocity of the model foot). *c*, An additional measurement obtained from the cavity-drag experiments was the time between impact and cavity seal (T_{seal}). T_{seal} was relatively invariant over a range of speeds (crosses in shaded

area). The step period (T_{step}) that the lizards use is typically shorter than T_{seal} , thereby allowing the lizards to lift their feet before the cavity collapses (circles). The foot model, based on detailed measurements of the foot of a 95-g lizard, was made of aluminum sheet metal (thickness = 0.762 mm). In measuring the lizard's foot, each toe was lightly compressed using finger pressure from above and the lateral toe fringe extended to obtain toe width. Also, the distance from the metacarpo-phalangeal joint (MP) to the tip of the toe and the distance from the heel to the MP was recorded for each toe. Finally, the width of the plantar surface at the level of the MP of the first toe was measured. Hydrodynamic measurements were obtained by attaching the foot model to a falling-missile apparatus which measured the upward water force on the foot while the model travelled directly downwards. The surface of the model foot remained parallel to the undisturbed water surface throughout the period of force measurement¹¹. The area S of the model foot ($0.610 \times 10^{-3} \text{ m}^2$) was measured by comparing the weight of the foot model with that of a standard of known area.



be neglected. We shall assume a lizard has only one foot in the water at a time, and the time both feet are out of the water is short or non-existent. Thus, the needed upward impulse is simply the product of a lizard's body weight (Mg) and the period between foot slaps (T_{step}),

$$\text{Needed impulse} = MgT_{\text{step}} \quad (3)$$

One way to test the simple hydrodynamic model is to determine whether the sum of the slap and stroke impulses equals the overall needed impulse. We found the sum of the upward impulses, on average, amounts to 102% of the needed impulse (triangle in Fig. 3). Also, we found that basilisks of this size class (~90 g) do not exploit the full range of slap and stroke impulse combinations. Rather, they use a narrow range of slap impulse contributions (12.6–22.6%), which is then supplemented with the stroke impulse (Fig. 3). The narrow range of impulse contributions is reflected in the stereotyped movements of these lizards as they run across the water surface. In contrast, smaller lizards have much more variable limb movements which may reflect a greater ability to adjust the relative contributions of the slap and stroke impulses⁷.

We also calculated the average mechanical power required to run on water (Fig. 3 legend). On average, the required power is approximately 29 W kg⁻¹. The power required is available: measurements using whole muscles from hindlimbs (*Dipsosaurus dorsalis*) indicate that lizard muscle can produce as much as 135 W kg⁻¹ at 35 °C (ref. 8). This suggests that at least 21% of the basilisk's body mass is involved in powering hindlimb motion.

Would humans be able to run on water? Our calculations indicate it would be impossible excluding artificial aids. An 80-kg subject who runs with a step period (T_{step}) similar to that of sprinters (0.26 s; ref. 9) would need to generate an upward impulse of 204 N s. To generate the support force, suppose the subject slaps the water surface and then strokes his feet (effective radius, 0.1 m) directly downwards to half his leg length (leg length, 1 m). Based on measurements of the drag forces produced by disks¹¹, we estimate the water-runner would need to stroke downwards through the water at almost 30 m s⁻¹, a velocity beyond human ability. Furthermore, to stroke downwards at this velocity, he would require an average power output almost 15 times greater than the maximum sustained power output for humans (20 W kg⁻¹ for humans running uphill)¹⁰. □

FIG. 3 Relative contributions of slap and stroke impulses to supporting the body weight on the water surface. To support their body weight, basilisk lizards must generate an upward force impulse that is equal to the product of their body weight and the period between steps (bold line). We found that, on average, the sum of the slap and stroke impulses is 102% of the overall needed impulse (triangle; slap/needed = 18.3 ± 2.72%; stroke/needed = 83.9 ± 12.8%; mean ± s.d.; 32 measurements). However, in a given step the total upward force impulse can either exceed (as high as 131%) or fall short (as low as 71.8%) of the needed impulse (open circles). A force deficit in a given step can be compensated by producing a force surplus in subsequent steps (filled circles; analyses of two steps taken with the same foot during a single sequence). These force impulse fluctuations can have a relatively small effect on the motion of the centre of mass because of the short time intervals (T_{step}) over which the accelerations due to the force surplus (or deficit) act. The clustering of the data points over a small range in slap-impulse contributions suggests that the lizards generate their support force in a stereotyped manner. Slap impulse was determined by multiplying the m_{virtual} (Fig. 2a) by the u_{peak} , typically occurring just after entry into the water, of the lizard's midfoot (equation (1)). Stroke impulse was calculated by integrating the upward component of Drag(t) with respect to time (equation (2)). Drag(t) is the product of foot area (S) and $C_D \cdot (0.5\rho u^2(t) + \rho gh(t))$; ρ is the mass density of water; C_D is the water-entry drag coefficient (Fig. 2b)¹¹. Step work was calculated by summing the work done on the fluid during both the impact and stroke phases. Average step power was obtained by dividing the step work by T_{step} (28.8 ± 5.60 W kg⁻¹; mean ± s.d.; 32 measurements).

2. Fish, F. E. *Physiol. Zool.* **55**, 180–189 (1982).
3. Harvald, Sv. Aa. *Resistance and Propulsion of Ships* (Krieger, Malabar, FL, 1991).
4. Rand, A. S. & Marx, H. *Copeia* **1**, 230–233 (1967).
5. Laerm, J. *Ecology* **55**, 404–411 (1974).
6. van Devender, R. W. in *Costa Rican Natural History* (ed. Janzen, D. H.) 379–380 (Univ. Chicago Press, Chicago, 1983).
7. Glasheen, J. W. thesis, Harvard Univ. (1995).
8. Johnson, T. P., Swoap, S. J., Bennett, A. F. & Josephson, R. K. *J. exp. Biol.* **174**, 199–213 (1993).
9. Armstrong, L. E., Costill, D. L. & Gehlsen, G. *Track Techn.* **87**, 2780–2782 (1984).
10. Kyle, C. R. & Caiozzo, V. J. *Eur. J. appl. Physiol.* **54**, 99–103 (1985).
11. Glasheen, J. W. & McMahon, T. A. *Physics of Fluids* (in the press).

ACKNOWLEDGEMENTS. We thank W. Meshaka and S. Wright for their advice and assistance. This study was supported by funds from the Office of Naval Research (to J.W.G.) and the Systems Development Foundation (to T.A.M.). Support for the expedition to Costa Rica was provided by Putnam and Tinker grants as well as logistical support from the Organization for Tropical Studies.

A role for stereoscopic depth cues in the rapid visual stabilization of the eyes

C. Busettini, G. S. Masson F. A. Miles*

Laboratory of Sensorimotor Research, National Eye Institute, National Institutes of Health, Bethesda, Maryland 20892, USA

PRIMATES have visual tracking systems that help stabilize the eyes on the surroundings by responding to retinal image motion at ultra-short latencies^{1,2}. However, as the observer moves through the environment, the image motion on the retina depends on the three-dimensional structure of the scene^{3,4}. We report here that the very earliest of these tracking responses is elicited only by objects moving in the immediate vicinity of the plane of fixation: objects nearer or farther are ignored. This selectivity is achieved by means of a stereoscopic depth mechanism which uses the fact that the two eyes have differing viewpoints, so only objects in the plane of fixation have images that occupy corresponding positions on the two retinæ. Such behaviour is readily explained by the known binocular properties of some motion-selective neurons in the visual cortex⁵. Some (stereanomalous) subjects showed highly specific tracking deficits as though lacking one subtype of these neurons.

When a passenger in a moving vehicle looks out at the passing scene and fixates on the far horizon, distant objects appear

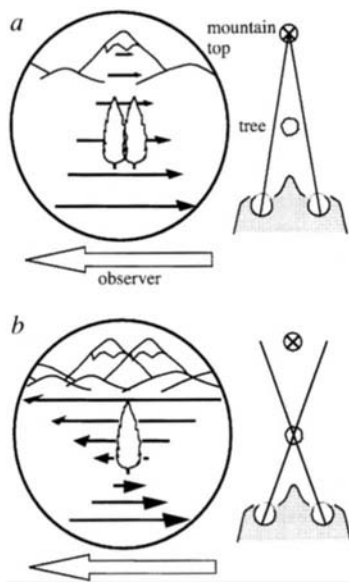


FIG. 1 The optic flow resulting from motion of the observer. At the left in *a* is a cartoon of the optic flow experienced by an observer moving to the left and fixating a distant object: retinal image motion is inversely proportional to the viewing distance. Given binocular viewing, the distant mountain in the plane of fixation is imaged in the same position on the two retinæ (shown in this cyclopean view as single), whereas the nearer tree is imaged in different positions (shown in this cyclopean view as double). The tree is said to have 'crossed disparity'. At the right in *a* is a plan view of the observer and the two objects: both eyes are aligned on the mountain so that from the right eye the tree is seen to the left of the mountain whereas from the left eye the tree is seen to the right of the mountain. In *b*, the observer fixates the tree, necessitating that he/she track to compensate for his/her own motion, thereby reversing the apparent motion of the more distant objects and creating a swirling pattern of optic flow. Now, the tree is imaged in the same position on the two retinæ (and hence shown single in the cyclopean view), whereas the mountain is imaged in different positions (and hence shown double in the cyclopean view). The mountain is said to have 'uncrossed disparity'. The dimensions of the eyes and their separation have been exaggerated to illustrate the disparity more clearly. In fact, disparity is much more evident with near viewing, which is also associated with the most vigorous optic flow and requires the most vigorous tracking from the observer to compensate for his own motion. For this reason, we used near viewing in our experiments.

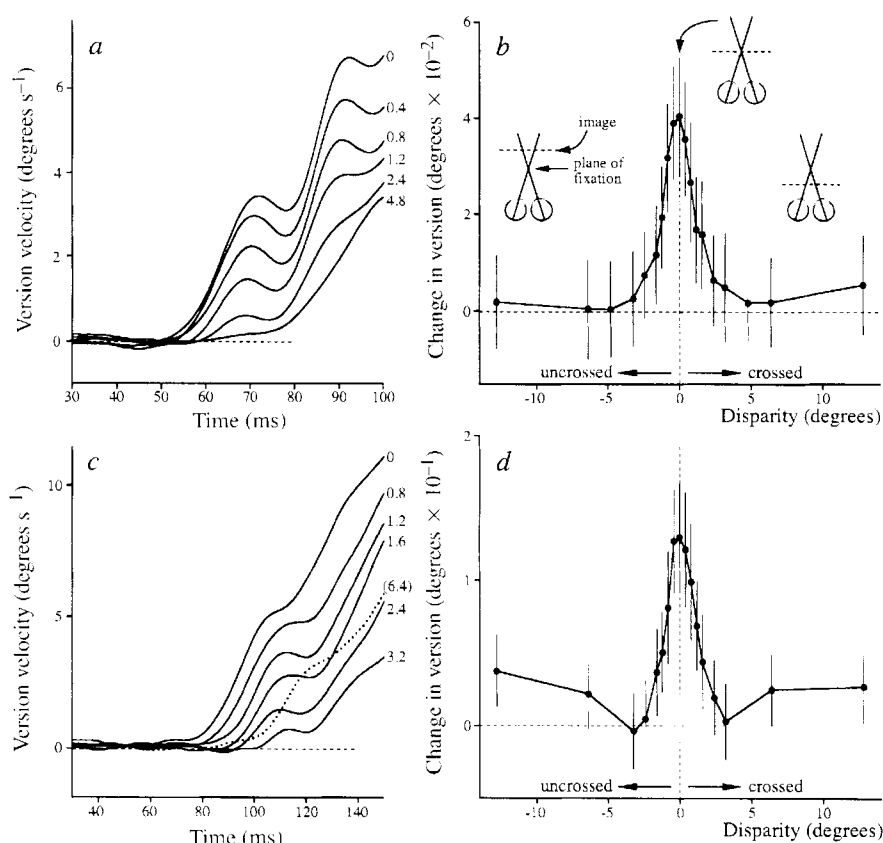
stationary whereas nearby ones seem to pass by rapidly as though the world were pivoting about the most distant point (Fig. 1*a*). If the passenger wishes to scrutinize objects in the middleground, then he/she must track them with his/her eyes, thereby compensating for the motion of the vehicle and shifting the pivot to the middleground; distant objects now appear to reverse their direction (Fig. 1*b*). We are here concerned with the visual mechanisms that allow an observer to track selectively the object in the middleground and ignore the competing motion (so-called motion parallax) of objects that are nearer or farther away. Humans can track the motion of images in the plane of fixation and ignore others outside the plane⁶. This behaviour might reflect high-level cognitive processing, but here we have investigated the suggestion that low-level visual motion detectors provide the initial drive to the tracking system and achieve this selectivity through a binocular stereomechanism⁶⁻⁸. This latter idea relies on the fact that the object on which the two eyes are aligned is imaged at corresponding positions on the two retinæ (the mountain in Fig. 1*a*; the tree in Fig. 1*b*), whereas objects that are nearer or farther away have images that occupy non-corresponding positions on the two retinæ (the tree in Fig. 1*a*; the mountain in Fig. 1*b*). The non-corresponding images are said to have 'binocular disparity'. We here describe the effects of such disparity on the tracking of large moving scenes by both humans and monkeys (ocular following).

As previously reported, the initial ocular following responses elicited by images moving in the plane of fixation were machine-like and had ultra-short latencies: less than 60 ms in monkeys and less than 85 ms in humans^{1,2}. These ocular following responses are of special interest because (1) they have many of the properties generally attributed to low-level motion detectors¹, and (2) they have been implicated in the visual stabilization of gaze during translational disturbances of the observer, which confront the system with conflicting motion signals such as those illustrated in Fig. 1*b*^{7,8}. The initial ocular following of all six subjects (3 monkeys, 3 humans) showed strong modulation with disparity, best seen from version velocity traces such as those shown for one monkey in Fig. 2*a* and for one human in Fig. 2*c*. In both of these cases, initial ocular following was greatest when the disparity was zero and decreased progressively as disparity increased. Initial responses were quantified by measuring the change in version during the initial rising phase of the responses and these measures are plotted in Fig. 2*b, d*. It is clear from these plots that the responses were maximal when the scene was imaged in the plane of fixation (zero

disparity) and decremented steeply with disparity, reaching minima with disparities of 3–5°, regardless of whether crossed or uncrossed. With larger disparities, the curves in Fig. 2*b* simply asymptoted near zero ('bell' profile), whereas those in Fig. 2*d* showed some recovery ('Mexican hat' profile, see also the dotted trace in Fig. 2*c*). With two notable exceptions, which are discussed below, all of the data obtained from all of the subjects had profiles between these extremes, with the humans tending slightly more towards the Mexican hat form and monkeys towards the bell form. Although not all of these curves peaked exactly at zero disparity, none deviated by more than 1°, and responses at zero disparity were always very close to the maximum. There were no clear differences between horizontal and vertical ocular following. Some disparity turning curves showed a slight asymmetry between the crossed and uncrossed stimuli. The lack of any obvious difference between the two species indicates that the monkey is a good animal model for the human.

Our data indicate that the motion detectors mediating the initial ocular following in both monkeys and humans are binocular and selectively sensitive to disparities close to zero, which under normal viewing conditions would make them responsive only to images in the immediate vicinity of the plane of fixation. It is known that there are neurons in the visual cortex (as early as V1) with binocular receptive fields that sense the relative positions of the images on the two retinæ and are therefore said to be 'disparity-selective'⁵. Some of these neurons, called 'tuned-disparity' neurons, respond to visual motion only over a restricted range of depths in ('tuned-zero') or close to ('tuned-near', 'tuned-far') the plane of fixation, and are also direction-selective⁵. Thus, they have exactly the properties required to mediate the present ocular following responses. However, the known range of depths covered by neurons of the 'tuned' type is smaller than the range of depths over which initial ocular following can be elicited⁵. We suggest that two important factors here are that our visual stimuli were much larger and moved at speeds that were much higher than those generally used in neurophysiological studies of stereomechanisms. The high speed of our stimuli would preclude a contribution from high-spatial-frequency components because of the temporal frequency ceiling, whereas the discrete, small size of the stimuli used in the neurophysiological experiments would mean that those studies lacked low-spatial-frequency components. It might be expected, therefore, that our responses would be mediated by neurons with coarser, more widely ranging, disparity-tuning curves than those reported. Apropos high-speed

FIG. 2 Dependence of ocular following on the horizontal disparity of the tracked images. *a*, Mean version velocity responses of monkey Lu in response to downward motion when images had crossed disparities (magnitude in degrees indicated at ends of traces). Note, version velocity is average velocity of both eyes. *b*, Plot of mean ($\pm$ s.d.) changes in version over time period 60–77 ms (starting from stimulus onset, downward motion) for monkey Lu. Each trace in *a* and each datum point in *b* is based on at least 165 responses. *c*, Mean version velocity responses of human subject G.S.M. in response to rightward motion when images had crossed disparities. *d*, Plot of mean ($\pm$ s.d.) changes in version over time period 85–118 ms (from stimulus onset, rightward motion) for human subject G.S.M. Each trace in *c* and each datum point in *d* is based on at least 66 responses. Positions of both eyes were recorded using the electromagnetic search coil technique¹⁰. Subjects faced a tangent screen (viewing distance 33 cm; subtense, $80 \times 80^\circ$) onto which two identical random-dot patterns were back-projected (dot diameters, $2''$; 50% density). Orthogonal polarizing filters in the projection paths ensured that each of the two eyes saw only one of the patterns, movements of which were produced by mirror galvanometers. One problem is that when the images seen by the two eyes are separated (that is, have disparity), compelling vergence eye movements are elicited that operate to realign the eyes on the two images and so eliminate the disparity¹¹. To avoid this, we used step-ramp motions of the two projected images and analysed only the very earliest ocular following responses that were generated open-loop, that is, before eye movements had had any chance to affect the visual stimuli on the retinae. The steps were disjunctive (horizontal disparities ranged 0 – 12.8°) and served to position the binocular image of the random-dot patterns in a new depth plane nearer ('crossed' disparities) or farther ('uncrossed' disparities) than the screen. The ramps started synchronously with the steps and were conjugate (both



images moved together at 40° s^{-1} , left, right, up, or down, with durations of 100 ms for monkeys and 200 ms for humans) and served to elicit ocular following by moving the patterns in the new depth plane. The magnitude and direction of the steps and ramps varied randomly from trial to trial, and catch trials were included in which no step-ramp was delivered.

stimuli, a recent psychophysical study⁹ showed that stereoacuity for periodic gratings is not degraded by speeds more than an order of magnitude greater than those that we have used here provided that the spatial frequency is low enough so that the temporal frequency does not exceed 30 Hz. Apropos large disparities, we have found that the ocular following responses elicited at short latency by monocular stimuli are much stronger than those elicited by binocular stimuli that have large disparities, indicating that the latter are subject to active suppression. Unfortunately, neurophysiological data are not available for large disparities such as those involved here.

That short-latency ocular following responses result from the activation of neurons that respond to motion only over a restricted range of disparities was reinforced by our finding that occasional tuning curves showed notches at a particular disparity. One such tuning curve taken from a monkey is shown in Fig. 3, and has a pronounced dip at zero disparity attributable to a specific deficiency in neurons of the 'tuned' type, especially the 'tuned-zero'. Interestingly, this stereoanomaly was limited to one direction of motion, rightward; in the other three directions the disparity tuning curves were normal, resembling those in Fig. 2. A similar observation (not shown) was made on one human subject, who exhibited a dip at zero disparity in the tuning curve for leftward ocular following, though this was much less pronounced than that in Fig. 3.

Together, these results demonstrate the existence of a machine-like visual tracking system that uses the stereoscopic properties of low-level motion detectors, such as have been described in the monkey's visual cortex, to help stabilize gaze on the plane of

fixation. It is possible that there are yet-to-be-discovered stereoscopic motion detectors that are tuned to depth planes well outside the plane of fixation and that the tracking system uses some form of selective attention to single out those detectors that are tuned to a particular depth plane. This would mean that, in our experiments, tracking was restricted to the plane of fixation merely

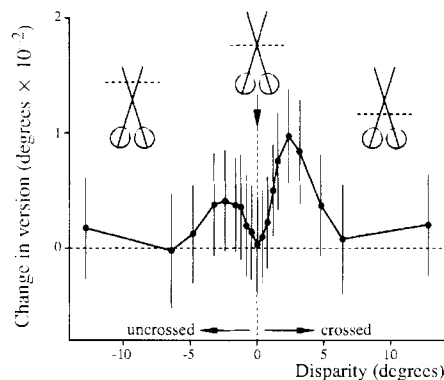


FIG. 3 Dependence of ocular following on the horizontal disparity of the tracked images: stereoanomalous profile. Mean ($\pm$ s.d.) data from monkey Sp in response to rightward motion. Each datum point is based on at least 185 responses. Conventions as for Fig. 2b, d.

because it happened to be the plane of attention. Of course, this need not always be the case in everyday life. □

Received 4 December 1995; accepted 26 January 1996.

1. Miles, F. A., Kawano, K. & Optican, L. M. *J. Neurophysiol.* **56**, 1321–1354 (1986).
2. Gellman, R. S., Carl, J. R. & Miles, F. A. *Visual Neurosci.* **5**, 107–122 (1990).
3. Gibson, J. J. *The Perception of the Visual World* (Houghton-Mifflin, Boston, 1950).
4. Busetini, C., Miles, F. A. & Schwarz, U. *J. Neurophysiol.* **66**, 865–878 (1991).
5. Poggio, G. F. *Cerebral Cortex* **5**, 193–204 (1995).
6. Howard, I. P. & Simpson, W. A. *Expl. Brain Res.* **78**, 309–314 (1989).
7. Miles, F. A., Schwarz, U. & Busetini, C. in *Representations of Vision: Trends and Tacit Assumptions in Vision Research* (ed. Gorea, A.) 185–199 (Cambridge Univ. Press, Cambridge, 1991).
8. Miles, F. A., Schwarz, U. & Busetini, C. in *The Head-Neck Sensory Motor System* (eds Berthoz, A., Vidal, P. P. & Graf, W.) 471–478 (Oxford Univ. Press, New York, 1992).
9. Morgan, M. J. & Castet, E. *Nature* **378**, 380–383 (1995).
10. Fuchs, A. F. & Robinson, D. A. *J. appl. Physiol.* **21**, 1068–1070 (1966).
11. Busetini, C., Krauzlis, R. J. & Miles, F. A. in *Contemporary Ocular Motor and Vestibular Research: A Tribute to David A. Robinson* (eds Fuchs, A. F., Brandt, T., Buttner, U. & Zee, D. S.) 312–319 (Thieme, Stuttgart, 1994).

ACKNOWLEDGEMENTS. Supported in part by the International Human Frontier Science Program. G.S.M. was supported by Le Ministère de la Recherche (France).

Role of microglia and host prion protein in neurotoxicity of a prion protein fragment

David R. Brown, Bernhard Schmidt* & Hans A. Kretzschmar†

Institut für Neuropathologie, * Zentrum Biochemie und Molekulare Zellbiologie, Biochemie II, Robert-Koch-Strasse 40, Universität Göttingen, 37075 Göttingen, Germany

The prion protein PrP^C is a glycoprotein of unknown function¹ normally found in neurons² and glia³. It is involved in diseases such as bovine spongiform encephalopathy (BSE), scrapie and Creutzfeldt–Jakob disease⁴. PrP^{Sc}, an altered isoform of PrP^C that is associated with disease, shows greater protease resistance and is part of the infectious agent, the prion^{5,6}. Prion diseases are characterized by neuronal degeneration, gliosis and accumulation of PrP^{Sc} (ref. 7). Mice devoid of PrP^C are resistant to scrapie⁸. A fragment of human PrP consisting of amino acids 106–126 that forms fibrils *in vitro* is toxic to cultured neurons^{9–11}. Here we show that this toxic effect requires the presence of microglia which respond to PrP106–126 by increasing their oxygen radical production. The combined direct and microglia-mediated effects of PrP106–126 are toxic to normal neurons but are insufficient to destroy neurons from mice not expressing PrP^C.

Synthetic PrP fragments have been used as a model to investigate neurodegeneration in prion diseases⁹. PrP106–126 is toxic to cortical cells cultured from embryonic-day-16 mice¹⁰. This toxic mechanism seems to require expression of PrP^C as cells from PrP^{0/0} mice¹², devoid of PrP^C, are resistant to the peptide's toxic effect¹⁰. PrP106–126 at 80 µM ($n = 6$) was also toxic to cultures from the cerebellum of 6-day-old normal mice (Table 1). PrP106–126 at 20 µM was not toxic in culture, with $97 \pm 4\%$ of cells surviving compared to untreated controls ($n = 5$). A control peptide having the PrP106–126 amino-acid sequence scrambled is not toxic in culture at 80 µM ($99 \pm 3\%$ survival, $n = 5$).

We co-cultured cerebellar cells with microglia or astrocytes from mouse cortex to investigate their contribution to the neurotoxicity of PrP106–126. Additional normal microglia enhanced PrP106–126 toxicity in cerebellar cell cultures, whereas microglia from PrP^{0/0} mice or normal astrocytes did not (Table 1). Further-

more, L-leucine methylester (LLME) was used to reduce significantly the number of microglia present in cerebellar cultures^{13,14}. LLME treatment of cerebellar cultures did not significantly reduce the overall survival of cells but abolished the susceptibility of the remaining cells to the toxic effect of PrP106–126, indicating that microglia are necessary for PrP106–126 toxicity.

If LLME treatment protects against the neurotoxicity of PrP106–126 by destroying microglia, then co-culturing LLME-treated cerebellar cultures with microglia should restore the toxic effect. Indeed, we found that co-culturing LLME-treated cells from normal mice with purified microglia from cerebral cortex of normal mice fully restored the toxicity of PrP106–126, and that co-cultures of microglia from PrP^{0/0} mice only partially restored it. This effect was not seen when LLME-treated cerebellar cells were co-cultured with astrocytes from normal mouse cortex, indicating that neurodestruction by PrP106–126 does not involve a direct effect on astrocytes.

Toxicity of 80 µM PrP106–126 increased proportionally to the log of the number of microglia added to the assay (Fig. 1). If the degree of cell destruction by PrP106–126 is related to the number of microglia, then a non-toxic concentration of peptide (20 µM) in the presence of increased numbers of microglia should also cause cell destruction. We co-cultured cerebellar cultures with microglia from normal or PrP^{0/0} mice and exposed them to 20 µM PrP106–126 for 10 days. The toxic effect caused by additional microglia (Fig. 2a) was weaker for microglia from PrP^{0/0} mice than for those from normal mice, suggesting that they are less responsive to PrP106–126.

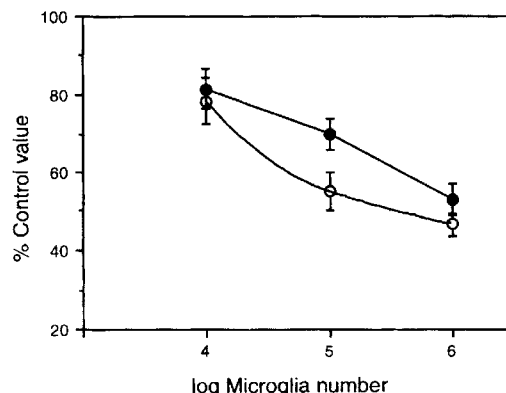


FIG. 1 LLME-treated normal cerebellar cells prepared on coverslips were co-cultured with microglia derived from normal (○) or PrP^{0/0} mice (●). Microglia were plated at cell densities varying from 10^4 to 10^6 cells per cm^2 in 24-well cloning trays. Co-cultured cells were treated with 80 µM PrP106–126. Coverslips were removed from wells after 10 d and an MTT assay done. Cell survival was determined as the per cent reduction in MTT assay absorbance compared with control cultures not treated with PrP106–126. The degree of cell destruction is proportional to the log of the number of microglia. The difference between the effect of normal and PrP^{0/0} microglia was significant only at a density of 10^5 cells per cm^2 . Values shown are means $\pm$ s.e. of 4–5 experiments with 3 determinations each. METHODS. Cerebellar cultures, survival assays and co-culturing are described in Table 1. Microglia were prepared from dissociating cerebral cortices¹³ of newborn normal or PrP^{0/0} mice. 6–7 cortices were treated with 0.5% trypsin and plated in 75- cm^2 culture flasks (Falcon) in DMEM supplemented with 10% FCS, 2 mM glutamine and 1% penicillin and streptomycin. Cultures were maintained at 37 °C with 10% CO_2 for 14 d until glial cultures were confluent. Microglia were isolated by shaking cultures at 260 r.p.m. for 2 h. Microglia dislodged into the medium were purified by preplating for 30 min in another flask. Contaminating cells were removed with the supernatant. Microglia were collected by further trypsinization and counted with a haemocytometer before finally plating at the number indicated. Astrocytes were isolated from the same glia preparations after most microglia had been removed by shaking. Further trypsinization of adherent cells and preplating for 30 min to remove remaining microglia gave pure astrocyte cultures.

† To whom correspondence should be addressed.

PrP106–126 was not toxic to cerebellar cells from PrP^{0/0} mice. The slightly but significantly (Student's *t*-test, $P < 0.05$) increased survival of PrP106–126-treated PrP^{0/0} cerebellar cells was not altered by co-culturing with normal microglia (Fig. 2b), indicating that the neurotoxicity of PrP106–126 depends on both specific interaction with PrP^C-expressing neurons and the presence of microglia.

We investigated the possibility that oxidants released by stimulated microglia might participate in the destruction of neurons^{13,15} and found that the antioxidants vitamin E and *N*-acetyl-cysteine could both block PrP106–126 toxicity in cerebellar cultures (Fig. 3a). Destructive oxidants such as superoxide radicals or nitric oxide are short-lived but can be assayed as they react to produce stable nitrite¹⁶. PrP106–126 caused an increase in nitrite in cultured microglia supernatants after 4 days (Fig. 3b); this effect was less pronounced in PrP^{0/0} microglia. These results support the hypothesis that PrP106–126-stimulated microglia kill neurons by oxidative damage.

If this oxidative stress caused by microglia were the only effect of PrP106–126, then PrP^{0/0} mouse cerebellar cells should be similarly destroyed. We treated microglia-reduced cerebellar cells with xanthine oxidase, an alternative source of superoxide radicals (Fig. 3c), and found that PrP^{0/0} mouse cells were more susceptible than normal cells to oxidative stress. However, PrP106–126 enhances cell death induced by xanthine oxidase only on LLME-treated cerebellar cells from normal mice. This strongly indicates that for PrP106–126 to exert a toxic effect on neurons, they must express PrP^C.

It is still debatable whether PrP106–126 is an appropriate model for the investigation of prion diseases. Although PrP106–126 does not appear to be infectious or cause conversion of PrP^C to PrP^{Sc}, it does cause apoptotic cell death in culture, as does PrP^{Sc} *in vitro*¹⁷ and *in vivo*¹⁸. Furthermore, our results indicate that

TABLE 1 Effect of PrP106–126 on survival of mixed cells from cultures of normal mouse cerebellum

Co-culture	Control	PrP106–126	LLME-treated	
			Control	PrP106–126
No co-culture	100 ± 2	55 ± 5	98 ± 2	93 ± 4
Normal microglia	98 ± 4	47 ± 4	98 ± 4	55 ± 5
PrP ^{0/0} microglia	97 ± 5	53 ± 4	97 ± 4	70 ± 4
Normal astrocytes	99 ± 3	58 ± 4	99 ± 3	100 ± 2

The effect of PrP106–126 on cerebellar cultures was determined by assessing survival of treated cells relative to untreated controls. PrP106–126 caused a significant cell loss (Student's *t*-test, $P < 0.05$). Co-culturing untreated cells with normal microglia significantly enhanced the toxicity of PrP106–126. Co-culturing cerebellar cells with PrP^{0/0} mouse microglia or normal astrocytes had no effect. L-leucine methylester (LLME) did not alter cell survival but blocked the toxic effect of the peptide. Co-culturing LLME-treated cerebellar cells with normal mouse microglia restored the toxicity of PrP106–126, co-culturing with PrP^{0/0} microglia gave partial restoration, and co-culturing with astrocytes had no effect. Values shown are means ± s.e. of at least 5 experiments with 3 determinations each. Cells were prepared by dissociating cerebella from 6-day-old mice produced by interbreeding C57BL/6J and 129/sv mice using 0.5% trypsin (Biochrom) in Hanks' medium (Biochrom). Cells were plated at $1-2 \times 10^6$ cells cm^{-2} in 24-well, poly-D-lysine ($50 \mu\text{g m}^{-2}$; Sigma) coated trays (Sarstedt). Cultures were maintained at 37 °C with 10% CO₂ in Dulbecco's minimal essential medium (Biochrom) containing 10% fetal calf serum, 2 mM glutamine and 1% penicillin and streptomycin. Peptides synthesized on a 9050 synthesizer (Millipore) were: PrP106–126, KTNMKHMGAAAAGAVVGGGLG; and 'scrambled' (random-order amino-acid residues of PrP106–126), NGAKALMGGHGATKVMVGAAA. Peptides at 80 μM (final) were added to cultures with fresh medium every two days. Cell survival was determined after 10 days in an MTT (Sigma)-based assay¹⁰. 5 mM L-leucine methyl ester (Sigma) was applied for 2 h to cultures after plating. Isolation of microglia and astrocytes is described in Fig. 1 legend. For co-culture, cerebellar cells were plated on coverslips and then transferred to cloning trays containing microglia or astrocytes. Coverslips were removed before MTT assay.

PrP106–126 interacts specifically with PrP^C-expressing cells. In addition, we show that PrP106–126 neurotoxicity in culture depends on the presence of and is induced by microglia. Oxidative stress is involved in this mechanism. Microglia are also activated in murine scrapie¹⁹ and there is evidence^{20,21} that microglia are

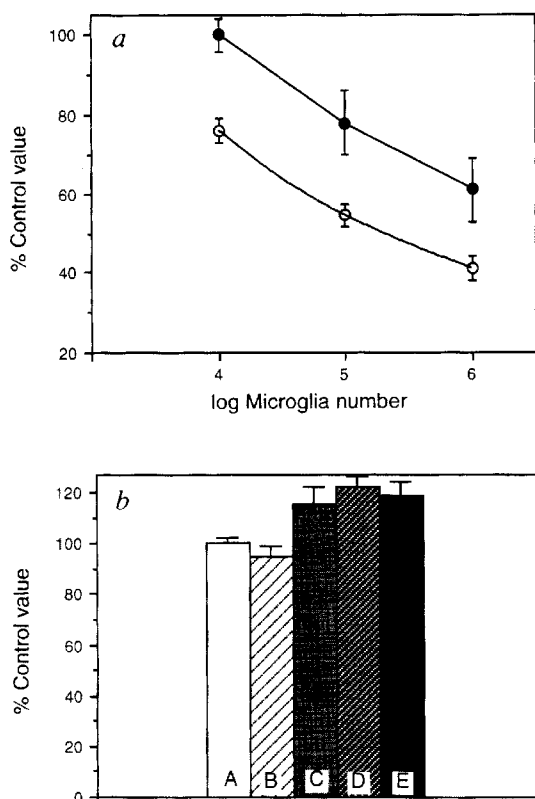


FIG. 2 The effect of co-culturing cerebellar cells from normal or PrP^{0/0} mice with high concentrations of microglia. a, Untreated normal cerebellar cells prepared on coverslips were co-cultured with microglia derived from normal (○) or PrP^{0/0} (●) mice. Microglia were plated at cell densities varying from 10^4 – 10^6 cells per cm^2 in 24-well trays. Co-cultured cells were treated with 20 μM PrP106–126, a concentration that is not toxic to cerebellar cells when cultured alone. Coverslips were removed from the wells after 10 days and an MTT assay done. Cell survival was determined as the per cent reduction in MTT assay absorbance compared with control cultures not treated with PrP106–126. The degree of cell destruction is proportional to the log of the number of microglia. Normal microglia were more effective in enhancing the toxic effect of the peptide than those from PrP^{0/0} mice, suggesting that these microglia are less sensitive to the effects of the peptide. b, Untreated PrP^{0/0} cerebellar cells prepared on coverslips were co-cultured with microglia from normal mice. Co-cultured cells were treated with 80 μM or 200 μM PrP106–126. Coverslips were removed from the wells after 10 days and an MTT assay done. Survival of co-cultured cerebellar cells (B) without PrP106–126 was similar to control cells (A) without co-culturing. Treatment with 80 μM PrP106–126 (C) caused a slight increase in cell survival above control which was unaltered (no significant difference, $P > 0.05$) by co-culturing cerebellar cells with microglia (D) even if the concentration of PrP106–126 was increased to 200 μM (E). This result indicates that expression of PrP^C is necessary for microglia to induce neuronal destruction in the presence of PrP106–126. Shown are means ± s.e. for 5 experiments with 3 determinations each.

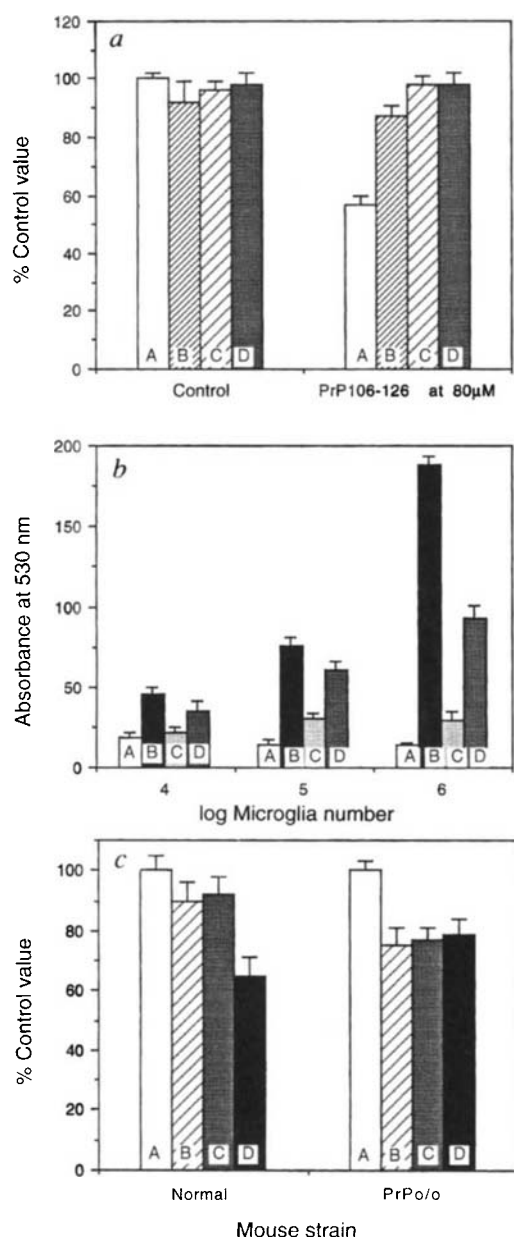


FIG. 3 Ability of PrP106-126 to produce oxidative stress in cell culture. *a*, As compared with control cultures (A), antioxidants vitamin E (60 mM in B; and 120 mM in C) or *N*-acetyl-cysteine (10 μM, D) blocked 80 μM PrP106-126 toxicity. Agents were applied every second day together with PrP106-126. Survival was determined on day 10 ($n = 5$). *b*, Nitrite production by microglia. Absorbance at 530 nm is proportional to nitrite concentration. Microglia exposed to 80 μM PrP106-126 (B, normal; D, PrP^{0/0}) gave significantly more nitrite than controls (A, normal; C, PrP^{0/0}). Enhanced nitrite production was proportional to the log of the number of microglia and was greater for normal microglia ($n = 12$). *c*, Normal or PrP^{0/0} cerebellar cells were treated with LLME to destroy microglia. Compared to controls (A) 5 μU of xanthine oxidase (a generator of oxygen radicals) was slightly toxic to normal LLME-treated cells but significantly more toxic (Student's *t*-test, $P < 0.05$) to PrP^{0/0} cerebellar cells, suggesting that the latter are less resistant to oxidative stress (B). 80 μM PrP106-126 alone did not destroy LLME-treated cerebellar cells (Table 1) and enhanced the effect of xanthine oxidase only on normal LLME-treated cells (D). 80 μM scrambled peptide applied with xanthine oxidase has no effect (C). PrP106-126 treatment enhances susceptibility of PrP^C-expressing cells to oxidative stress ($n = 4$, 2–3 determinations each). **METHODS.** As for Table 1 and Fig. 1. The nitrite assay has been described²². Microglia placed in culture at 10^4 to 10^6 cells per cm^2 were exposed to 80 μM PrP106-126 for 4 days; 0.5 ml medium was used for measurements. Xanthine oxidase (Boehringer) was applied with 5 μM xanthine for 24 h. 80 μM PrP106-126 was applied to xanthine oxidase-treated cultures at the same time. Values shown are means $\pm$ s.e.

stimulated by β -amyloid to produce neurotoxic agents. Microglia appear to be an important mediator of neuronal death in degenerative diseases of the brain. □

Received 5 October 1995; accepted 31 January 1996.

1. Herms, J. W., Kretschmar, H. A., Titz, S. & Keller, B. U. *Eur. J. Neurosci.* **7**, 2508–2512 (1995).
2. Kretschmar, H. A., Prusiner, S. B., Stowring, L. E. & DeArmond, S. J. *Am. J. Path.* **122**, 1–5 (1986).
3. Moser, M., Colello, R. J., Pott, U. & Oesch, B. *Neuron* **14**, 509–517 (1995).
4. Prusiner, S. B. *Science* **252**, 1515–1522 (1991).
5. Bolton, D. C., McKinley, M. P. & Prusiner, S. B. *Science* **218**, 1309–1311 (1982).
6. Caughey, B. & Raymond, G. J. *J. Biol. Chem.* **266**, 18217–18223 (1991).
7. DeArmond, S. J. *et al. Neurology* **37**, 1271–1280 (1987).
8. Büeler, H. *et al. Cell* **73**, 1339–1347 (1993).
9. Forloni, G. *et al. Nature* **362**, 543–546 (1993).
10. Brown, D. R., Herms, J. & Kretschmar, H. A. *Neuroreport* **5**, 2057–2060 (1994).
11. Selvaggi, C. *et al. Biochem. biophys. Res. Commun.* **194**, 1380–1386 (1993).
12. Büeler, H. *et al. Nature* **356**, 577–582 (1992).
13. Giulian, D. & Baker, T. J. *J. Neurosci.* **6**, 2163–2178 (1986).
14. Giulian, D. *et al. J. Neurosci.* **8**, 709–714 (1988).
15. Betz Corradin, S. *et al. Glia* **7**, 255–262 (1993).
16. Beckman, J. S. *et al. Proc. natn. Acad. Sci. U.S.A.* **87**, 1620–1624 (1990).
17. Müller, W. E. G. *et al. Eur. J. Pharmac.* **246**, 261–267 (1993).
18. Giese, A., Groschup, M., Hess, B. & Kretschmar, H. A. *Brain Path.* **5**, 213–221 (1995).
19. Williams, A. E., Lawson, L. J., Perry, V. H. & Fraser, H. *Neuropath. appl. Neurobiol.* **20**, 47–55 (1994).
20. Meda, L. *et al. Nature* **374**, 647–650 (1995).
21. Frackowiak, J. *et al. Acta neuropath.* **84**, 225–233 (1992).
22. Elstner, E. F. & Heupel, A. *Ann. Biochem.* **70**, 616–620 (1976).

ACKNOWLEDGEMENTS. We thank C. Weissmann for PrP^{0/0} mice. This study was supported by research grants from the Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie and the Wilhelm-Sander-Stiftung.

Functional contributions of $\alpha 5$ subunit to neuronal acetylcholine receptor channels

J. Ramirez-Latorre, C. R. Yu, X. Qu, F. Perin*, A. Karlin† & L. Role‡

Center for Neurobiology and Behavior, Department of Anatomy & Cell Biology, * Ecole Normale Supérieure, † Center for Molecular Recognition, College of Physicians and Surgeons of Columbia University, P.I. Annex 807, 722 168th Street, New York, New York 10032, USA

LIGAND-GATED ion channels are multi-subunit complexes where each subunit-type is encoded by several related genes. Heterologous expression of any one of the neuronal nicotinic acetylcholine receptors (nAChR) α -type subunits, either alone or with any β -type subunit, typically yields functional nAChR channels^{1–5}. A striking exception is the nAChR $\alpha 5$ subunit: although apparently complexed with $\beta 2$ and $\beta 4$ nAChR subunits in neurons^{6–8}, and expressed in a subset of neurons within the central and peripheral nervous systems^{9,10}, heterologous expression of $\alpha 5$, either alone or with any β -type subunit has failed to yield functional channels^{1,11}. We demonstrate here that $\alpha 5$ does participate in nAChRs expressed in heterologous systems and in primary neurons, and further that $\alpha 5$ contributes to the lining of functionally unique nAChR channels, but only if coexpressed with both another α - and β -type subunit. Furthermore, channels containing the $\alpha 5$ subunit are potentially activated and desensitized by nanomolar concentrations of nicotine.

The neuronal nicotinic receptors are ligand-gated, cation-selective channels, made of combinations of at least two subunit types^{3–5,12–14}. Three of the nAChR subunit gene products, $\alpha 5$, $\alpha 6$ and $\beta 3$, have failed to yield agonist-gated currents when

‡ To whom correspondence should be addressed.

coexpressed in oocytes with any one of the other α or β genes^{1,9} (M. Ballivet, personal communication). The apparent lack of function in these studies has been particularly beguiling because $\alpha 5$ messenger RNA is expressed in specific regions of mammalian central nervous system (CNS) and coimmunoprecipitated with $\beta 2$ and $\beta 4$ (ref. 8) from avian nerve tissue extracts.

Injection of chick $\alpha 5$ mRNA alone or coinjection of $\alpha 5$ with either $\alpha 4$ or $\beta 2$ yielded no detectable macroscopic currents in response to acetylcholine (ACh) or nicotine (>1 mM; Fig. 1a). In contrast, coinjection of all three mRNA species yielded ACh-induced macroscopic currents. When similar quantities of mRNA were injected, the maximum current obtained from $\alpha 5 + \alpha 4 + \beta 2$ was typically greater than that obtained from $\alpha 4 + \beta 2$ (Fig. 1b). Repeated application of 100 μ M ACh evokes currents of consistent amplitude in oocytes coexpressing either $\alpha 4 + \beta 2$ or $\alpha 5 + \alpha 4 + \beta 2$. The same test protocol using nicotine as the agonist reveals a greatly enhanced inactivation rate of the $\alpha 5 + \alpha 4 + \beta 2$ currents as compared with those obtained by injection of $\alpha 4 + \beta 2$ (Fig. 1c).

After coexpression of $\alpha 4 + \beta 2$, the dependence of current on agonist concentration was fitted by the Hill equation with a half-maximal effector concentration (EC_{50}) of 0.8 μ M both for ACh and for nicotine. In contrast, after coexpression of $\alpha 5 + \alpha 4 + \beta 2$, the dose-response data were fitted best by the sum of two Hill

equations, one with a low EC_{50} , corresponding to that of $\alpha 4 + \beta 2$, and one with an EC_{50} roughly 125-fold greater (Fig. 2). For ACh, the two EC_{50} values were 0.8 μ M and 100 μ M. The EC_{50} for nicotine was increased roughly 15-fold by the coexpression of $\alpha 5 + \alpha 4 + \beta 2$.

In both outside-out and cell-attached patches, channels expressed after the injection of $\alpha 4 + \beta 2$ mRNA had a conductance of 24 ± 3 pS ($n = 10$; Fig. 3a, c), in agreement with previous reports^{2,15}. Coinjection of $\alpha 5 + \alpha 4 + \beta 2$ yielded an additional conductance class of 44 ± 4 pS ($n = 11$). The larger conductance class was detected with ratios of $\alpha 5 : \alpha 4 : \beta 2$ mRNA as low as 1 : 1 : 1, and its relative abundance increased with the injection of increased amounts of $\alpha 5$ relative to $\alpha 4$ and $\beta 2$ mRNA (Fig. 3b, d). Because the 44 pS channel was never detected in the absence of $\alpha 5$, this channel is dependent on the coinjection of $\alpha 5$ with $\alpha 4$ and $\beta 2$. Although the conductances of nAChRs composed of $\alpha 4 + \beta 2$ and $\alpha 5 + \alpha 4 + \beta 2$ were identical whether gated by nicotine or by ACh (Fig. 3c, d), the opening probability and the rate of desensitization of nicotine-gated $\alpha 5 + \alpha 4 + \beta 2$ channels were greater than those of $\alpha 4 + \beta 2$, even at concentrations of nicotine <1 μ M (Fig. 3c, d, right panel).

Examination of native nAChR confirms that $\alpha 5$ may contribute to the biophysical and pharmacological profile of nACh channels *in vivo*, where $\alpha 5$ is predominantly expressed in limbic and

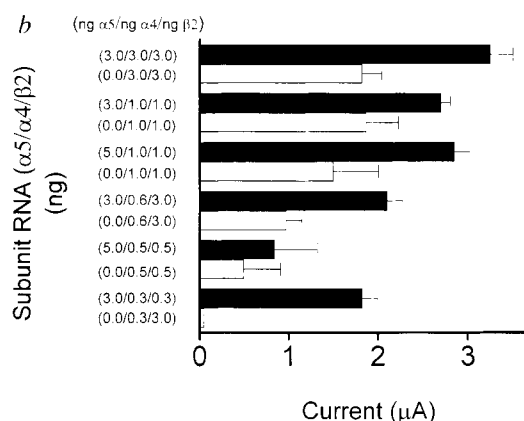
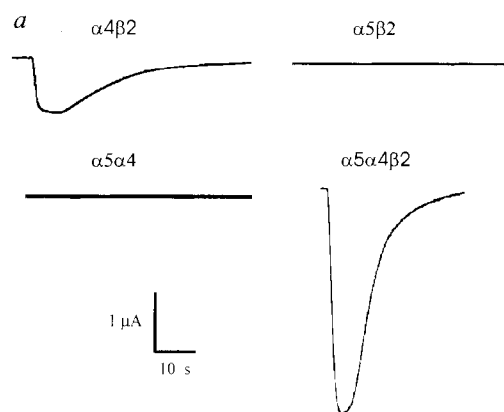
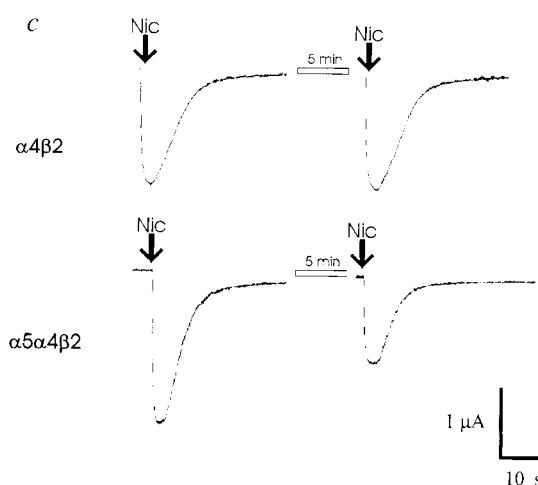


FIG. 1 Coexpression of $\alpha 5$ with $\alpha 4 + \beta 2$ yields ACh-evoked macroscopic currents larger than those with $\alpha 4 + \beta 2$ alone. a, ACh-evoked currents in oocytes and injection of mRNA as follows: (proceeding clockwise, beginning with upper left current) $\alpha 4 : \beta 2$ (1 ng : 1 ng), $\alpha 5 : \beta 2$ (3 ng : 3 ng), $\alpha 5 : \alpha 4$ (3 ng : 3 ng) and $\alpha 5 : \alpha 4 : \beta 2$ (5 ng : 1 ng : 1 ng). Robust currents are seen following coexpression of $\alpha 4 + \beta 2$ or $\alpha 5 + \alpha 4 + \beta 2$, whereas no current is detected with application of ACh (100 μ M–10 mM) in oocytes previously injected with either $\alpha 5 + \beta 2$ or $\alpha 5 + \alpha 4$. b, Data summary of ACh-evoked macroscopic current amplitudes in oocytes previously injected with varying amounts of $\alpha 4 + \beta 2$ (open bars) versus $\alpha 5 + \alpha 4 + \beta 2$ (black bars) mRNA. The numbers along the ordinate axis indicate the amount (in ng) of $\alpha 5$, $\alpha 4$ and $\beta 2$ RNA injected (for example, 3.0/0.6/3.0 indicates coinjection of 3 ng $\alpha 5$, 0.6 ng $\alpha 4$ and 3 ng $\beta 2$). c, Nicotine evoked currents in oocytes and injection of RNA as follows: upper panel, $\alpha 4 : \beta 2$ (1 ng : 1 ng); lower panel, $\alpha 5 : \alpha 4 : \beta 2$ (10 ng : 1 ng : 1 ng). Repeated Nicotine application (100 μ M, 5-s duration, 5-min intervals) evokes currents of consistent amplitude in oocytes injected with $\alpha 4/\beta 2$ complementary RNA (upper panel). The same application causes a marked desensitization of the currents evoked in $\alpha 5/\alpha 4/\beta 2$ -injected oocytes (lower panel).

METHODS. Physiology: Macroscopic currents were recorded with a 500 Gene-Clamp voltage-clamp (Axon Instruments) with an active ground. The voltage, current and active ground electrodes were filled with 3 M KCl such that voltage and current recording electrodes were ≈ 1 –5 M Ω . $V_{\text{clamp}} = -60$ mV. Extracellular recording solution (mM): 82.5 NaCl, 2 KCl, 1 MgCl₂, 10 HEPES, 1 μ M atropine, pH 7.5. ACh concentration, 10 μ M for $\alpha 4/\beta 2$ and 1,000 μ M (10 times the EC_{50} for both) for $\alpha 5/\alpha 4/\beta 2$. Oocyte and RNA preparation: Oocytes were obtained following incubation of *Xenopus* ovaries in 2 mg ml⁻¹ collagenase (Type I, Sigma) in ND96 (Specialty Media). The oocytes were incubated at room temperature for 3–4 h, and



the connective tissue removed. The eggs were then washed 4 times in Barth's media, transferred to L-15, incubated overnight at 18 °C, and the RNA injected the next day. RNA was prepared from cDNAs in PGH19, a vector which contains 5' and 3' untranslated *Xenopus* globin regions. This vector was linearized and used as a template for run-off transcription, using the T7 promoter. The RNA used for transcription was 0.2 μ g μ l⁻¹, and the volume injected was 20 nl per oocyte.

autonomic regions⁶⁻⁹. Subunit-specific antisense oligonucleotides were used to functionally 'delete' $\alpha 5$ from sympathetic neurons, where the nAChR channel profile and contribution of other nAChR subunits have been characterized in previous studies^{16,17}. Treatment with control oligonucleotides (see legend to Fig. 3) has no effect on the three nAChR subtypes normally expressed (≈ 15 , 33 and 53 pS). Treatment with $\alpha 5$ -specific antisense oligonucleotides and assay by single-channel recording, reveals a selective deletion of the largest (53 pS) conductance channel and no alteration in expression of either the 15 or the 33 pS nAChRs subtypes (Fig. 3e, f). Furthermore, assay of the agonist dependence of the residual current in these ' $\alpha 5$ -minus' neurons reveals significant decreases in the EC_{50} values (C.R.Y. and L.R., manuscript in preparation). Thus the properties of nAChRs deleted by $\alpha 5$ -antisense treatment are reminiscent of heterologously expressed nAChRs that include $\alpha 5$.

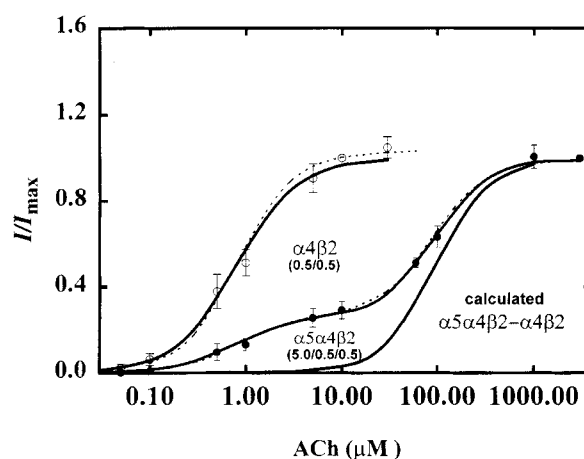
Although these results imply the participation of $\alpha 5$ in native and heterologously expressed nAChR complexes, alternative possibilities include an influence of $\alpha 5$ on the assembly of nAChRs with new $\alpha : \beta$ stoichiometries. We sought direct proof for participation of the $\alpha 5$ subunit *per se* in nAChR channel complexes. If the $\alpha 5$ subunit participates directly in a functional complex with $\alpha 4$ and $\beta 2$, then a region of the $\alpha 5$ sequence is likely to contribute to the ion-conduction pathway. To test this, we mutated individual residues in the M2-membrane-spanning segment (Fig. 4b, inset) of the $\alpha 5$ sequence to cysteine and expressed the mutant $\alpha 5$ mRNA together with $\alpha 4$ and $\beta 2$ mRNA. The residues mutated align with those previously shown to be exposed in the channel of the mouse muscle AChR¹⁸. In mouse muscle AChR, these residues, when mutated to cysteine, reacted with charged, hydrophilic, sulphhydryl-specific methane thiosulphonate (MTS) derivatives, and this reaction inhibited the ACh-induced current. We injected oocytes with wild-type or mutant $\alpha 5$ and $\alpha 4 + \beta 2$, where the mutated $\alpha 5$ sequences were altered as follows: S240C, L246C, V247C, S248C, L254C and E257C (Fig. 4b). We tested the susceptibility of these mutants to the positively charged MTS derivative, MTS-ethyltrimethylammonium¹⁸ (MTSET; Fig. 4a).

MTSET (1 mM for 2 min) irreversibly inhibited ACh-induced currents when the nAChRs contained the mutants, $\alpha 5$ -S240C, $\alpha 5$ -S248C, $\alpha 5$ -L254C and $\alpha 5$ -E257C. The extent of irreversible inhibition of ACh-evoked currents (100 μ M ACh) averaged 34%, 21%, 32%, and 28%, respectively (Fig. 4b). MTSET did not irreversibly alter ACh-evoked currents when nAChR complexes contained wild-type $\alpha 5$ or the $\alpha 5$ mutants, L246C or V247C; that is, as in mouse muscle AChR, only certain of the residues of M2 are accessible to MTSET in the channel (Fig. 4b). The inhibition of cysteine-substituted mutants of $\alpha 5$ demonstrates that $\alpha 5$ is part of a functional receptor complex and that the $\alpha 5$ -M2 segment contributes to the lining of the channel.

The proposed contributions of $\alpha 5$ to ligand-binding and nAChR conductance are compatible with previous studies. For example, in muscle nicotinic receptor, the aligned residues in α , β , γ and δ at P2 (Fig. 4b) are major determinants of cationic selectivity and of ionic conductance¹⁹. The conductance is determined by the volume and hydrophobicity of the substituted side chain²⁰. The participation of $\alpha 5$ in a pentameric complex would imply a change of at least one $\alpha 5$ subunit for either an $\alpha 4$ subunit or a $\beta 2$ subunit. In either case this would yield a substitution of a Thr by a Ser (P2), with a resultant decrease in side-chain volume. Also, an increase in the negative charges at P20, (the 'outer ring' of charges) can alter conductance¹⁹. The $\alpha 5$ subunit contains Glu Glu at P19 and P20, compared to Thr Glu in $\alpha 4$, and Ser Lys in $\beta 2$ (Fig. 4b). Together these three differences might account for the increase in conductance. The reaction of E257C at P19 with MTSET shows that this residue is exposed in the channel.

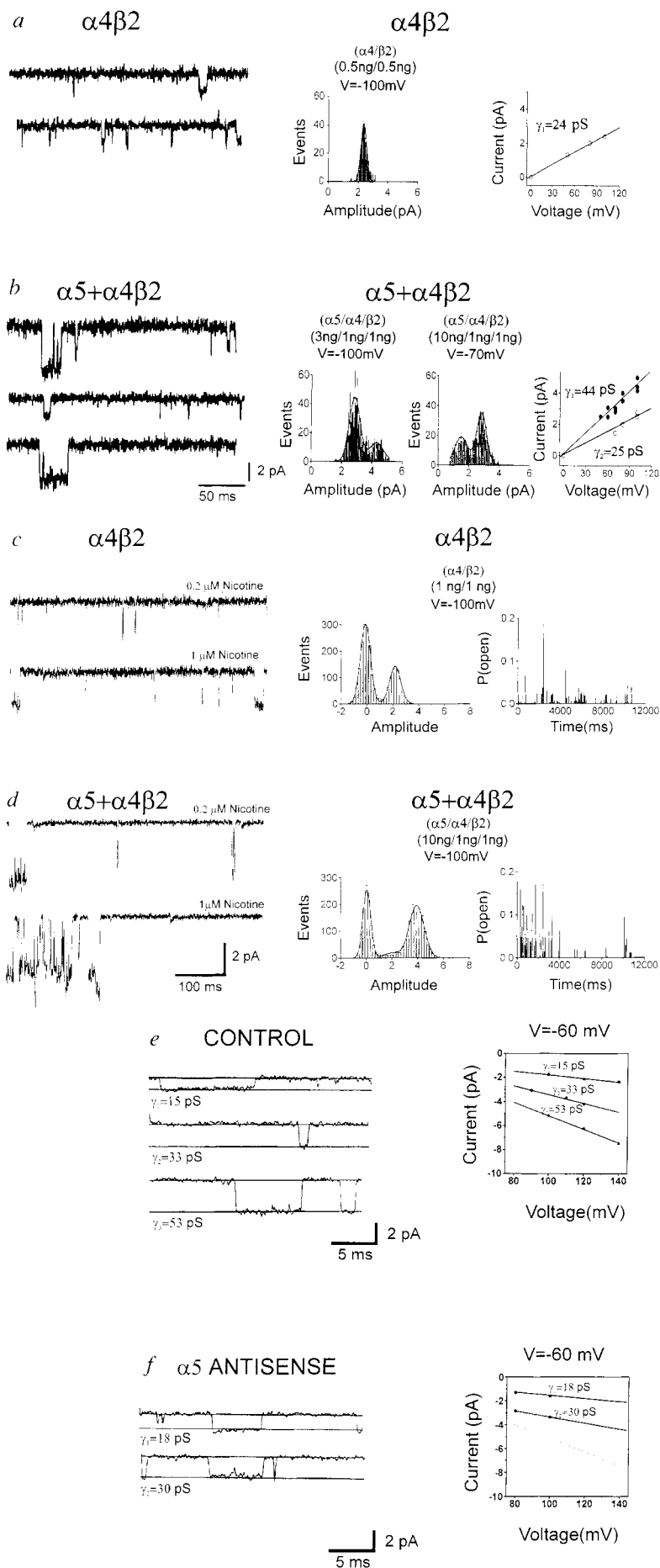
Six residues in the Torpedo α -subunit have been labelled by ACh-binding-site affinity reactions (see ref. 21 for review), and these residues are highly conserved among α -subunits. One of these residues is α Tyr 190^{22,23}, the mutation of which has profound effects on ACh binding and gating²⁴⁻²⁶. Only in $\alpha 5$, where it is replaced with an Asp, is Tyr 190 not conserved. This alteration could explain the 125-fold increase in EC_{50} of a population of receptors, assuming that the residues around Asp190 in $\alpha 5$, by analogy with $\alpha 1$, participate directly in the binding of ACh. Another possibility is that another region of $\alpha 5$, by analogy with

FIG. 2 The coinjection of $\alpha 5$ with $\alpha 4 + \beta 2$ cRNA into *Xenopus* oocytes yields a second, lower apparent affinity component to the ACh dose-response curve. Concentration-dependence of ACh-evoked macroscopic currents in oocytes previously injected with $\alpha 4 + \beta 2$ (open circles; 0.5 ng $\alpha 4$, 0.5 ng $\beta 2$) compared with those injected with $\alpha 5 + \alpha 4 + \beta 2$ (filled circles; 5 ng $\alpha 5$, 0.5 ng $\alpha 4$, 0.5 ng $\beta 2$). ACh-evoked responses with coinjection of $\alpha 4 + \beta 2$ RNA are fitted by the function $f(x) = (1 + (K_1/x)^{n_1})^{-1}$ with $K_1 = 0.8 \pm 0.07 \mu$ M, $n_1 = 1.34 \pm 0.13$, where K_1 indicates a single $EC_{50}^1 = 0.8 \mu$ M. In contrast, coinjection of $\alpha 5$ with $\alpha 4 + \beta 2$ yields a biphasic dose-response curve fitted by the function $g(x) = (A - 1)(1 + (K_1/x)^{n_1})^{-1} + A(1 + (K_2/x)^{n_2})^{-1}$. Using a self-consistent iterative procedure to fit these parameters, we obtain $A = 0.80 \pm 0.01$, $K_2 = 101.4 \pm 3 \mu$ M, $n_2 = 1.43 \pm 0.1$. The third line represents the calculated contribution of the low-affinity curve alone, $h(x) = (1 + (K_2/x)^{n_2})^{-1}$. $K_2 = EC_{50}^2$. This analysis indicates that the high-affinity component of the biphasic curve represents the $\alpha 4\beta 2$ contribution to the current, and the value of A indicates that $\sim 20\%$ of the current is carried by $\alpha 4\beta 2$ (at this ratio of $\alpha 5$, $\alpha 4$ and $\beta 2$ subunit RNAs 10 : 1 : 1). The shift between the high- and low-affinity EC_{50} is ~ 125 fold. Increasing the amount of $\alpha 5$ RNA relative to the $\alpha 4$ and $\beta 2$ (for example to $\alpha 5$ (10 ng) $\alpha 4$ (0.5 ng) $\beta 2$ (0.5 ng)) does not further decrease the contribution of the $\alpha 4\beta 2$ current, such that a $\alpha 5 : \alpha 4 : \beta 2$ ratio of 0 : 0.5 : 0.5 results in $EC_{50}^1 = 0.8 \mu$ M, $EC_{50}^2 = 0$, $A = 0$ ($n = 9$). A ratio of 5 : 1 : 1 produces $EC_{50}^1 = 0.75 \mu$ M, $EC_{50}^2 = 96.4 \mu$ M, $A = 0.27$ ($n = 3$). A ratio of 10 : 1 : 1 results in $EC_{50}^1 = 0.83 \mu$ M, $EC_{50}^2 = 101.4 \mu$ M, $A = 0.2$ ($n = 6$) and a ratio of 20 : 1 : 1, $EC_{50}^1 = 0.7 \mu$ M, $EC_{50}^2 = 100 \mu$ M and $A = 0.2$ ($n = 3$). The analysis using the Hill equation assumes infinite cooperativity, and in spite of its widespread use, is valid only in the limiting case. Using a more realistic cooperative binding scheme, one obtains for the case of two binding sites with an energy of interaction w_{12} , the grand canonical partition function, $\xi = 1 + 2x + yx^2$, where $x = c/K$ and $y = \exp(-w_{12}/KT)$ (c is the ligand concentration K is Boltzmann constant and T is the temperature). For the occupancy function $\theta = (x/2)(\delta \ln \xi / \delta \ln x) = (x + yx^2) /$



$(1 + 2x + yx^2)$. Using this function we obtain, $K_1 = 2.5 \mu$ M, $K_2 = 280 \mu$ M, $y = 10$ ($w_{12} = -1.347$ cal mol⁻¹ at 300 K). The fit of this functions to the $\alpha 4\beta 2$ and $\alpha 5\alpha 4\beta 2$ dose-response curves is shown as dotted lines in the figure. The equivalent Hill coefficient is related to fluctuations in the extent of binding and is obtained by $n = 4(\delta \ln \theta / \delta \ln x)_{\theta=1/2}$, which is in this case, $n = 2y^{1/2} / (1 + y^{1/2})$, and gives $n = 1.57$ for both the low- and high-affinity curves. Because of the size of the oocyte, there is considerable spread of the activation of receptors, which distorts the measurement of the macroscopic current kinetics. This measurement is further complicated by inaccessibility of up to 30% of receptors because of membrane infoldings and the presence of the vitelline membrane (unpublished observations).

FIG. 3 Coexpression of $\alpha 5$ with $\alpha 4 + \beta 2$ yields a new nAChR channel subtype of higher conductance. ACh-gated single-channel currents in outside-out patches from oocytes previously injected with $\alpha 4 + \beta 2$ (a) compared with those obtained following the coinjection of $\alpha 5 + \alpha 4 + \beta 2$ (b) RNA injection. a, Left, single channels recorded from an outside-out patch in an oocyte injected with $\alpha 4$ (1 ng) $\beta 2$ (1 ng) RNA. Membrane potential -100 mV. ACh concentration, $1 \mu\text{M}$. Middle, a single gaussian fit yields $m_1 = 2.36 \pm 0.4$ pA. Right, the single-channel conductance from linear regression of the I/V plot is 23.96 ± 3 pS. b, Left, typical single-channel records from an outside-out patch following injection of $\alpha 5$ (3 ng) $\alpha 4$ (1 ng) $\beta 2$ (1 ng) RNA in oocyte. Membrane potential -100 mV. ACh concentration, $100 \mu\text{M}$. Middle, the amplitude histogram of the recorded events is fit by the sum of two gaussian curves with $m_1 = 2.8 \pm 0.7$, $m_2 = 4.3 \pm 0.9$. Increasing the amount of $\alpha 5$ mRNA relative to $\alpha 4 + \beta 2$ decreases the number of events within the m_1 gaussian and increases the number of events described by m_2 . Right, the single channel conductance of the higher amplitude class is 44 ± 3.6 pS as estimated by linear regression of the I/V plot. Coinjection of $\alpha 5$ with $\alpha 4 + \beta 2$ is without effect on the conductance of the smaller channel (25 ± 3.5 pS). The linear regressions to assess single-channel conductance were constrained to pass through the origin (a, b). c, Left, single channels recorded from an oocyte injected with $\alpha 4$ (1 ng) $\beta 2$ (1 ng) RNA. Membrane potential, -100 mV. Nicotine concentration, $0.2 \mu\text{M}$ or $1 \mu\text{M}$, as indicated. Middle, amplitude histogram of the recorded events is fit by a single gaussian (excluding zero events) with mean and s.d. $m_1 = 2.3 \pm 0.8$ pA. A slope conductance measurement for this patch, using events at -60 , -80 and -100 mV yields $\gamma = 21 \pm 4$ pS. d, Left, typical single channel records from an outside-out patch from an oocyte injected with $\alpha 5$ (1 ng) $\alpha 4$ (0.1 ng) $\beta 2$ (0.1 ng). Nicotine concentration $0.2 \mu\text{M}$ or $1 \mu\text{M}$ as indicated. Middle, amplitude histogram, fit by two gaussians (excluding zero events) with means and s.d., $m_1 = 2.1 \pm 1.6$ pA and $m_2 = 3.91 \pm 1.1$ pA. Slope conductance $\gamma = 45 \pm 6$ pS (using events at -60 , -80 and -100 mV). e, Typical single-channel recordings from chick sympathetic neurons (ED11) reveals three conductance classes ($\gamma_1 = 15$ pS, $\gamma_2 = 33$ pS, $\gamma_3 = 53$ pS). f, Treatment with $\alpha 5$ antisense deletes the high conductance class ($\gamma_3 = 53$ pS) and is without effect on lower conductance classes ($\gamma_1 = 18$ pS, $\gamma_2 = 30$ pS). Neurons were dispersed from ED11 lumbar sympathetic ganglia, maintained *in vitro* and treated with antisense as described previously²⁷. A three-base mismatch $\alpha 5$ -antisense sequence served as control. Single-channel recordings were obtained using standard techniques²⁸ with an Axon 200A patch-clamp (Axon Instruments). Solutions (in mM): Solution 1 (bath): 115 NaCl, 2 KCl, 1 CaCl_2 , 1 MgCl_2 , 10 HEPES, pH 7.5; solution 2 (pipette): 115 KCl, 2 NaCl, 0.2 EGTA, 10 HEPES, pH 7.5.



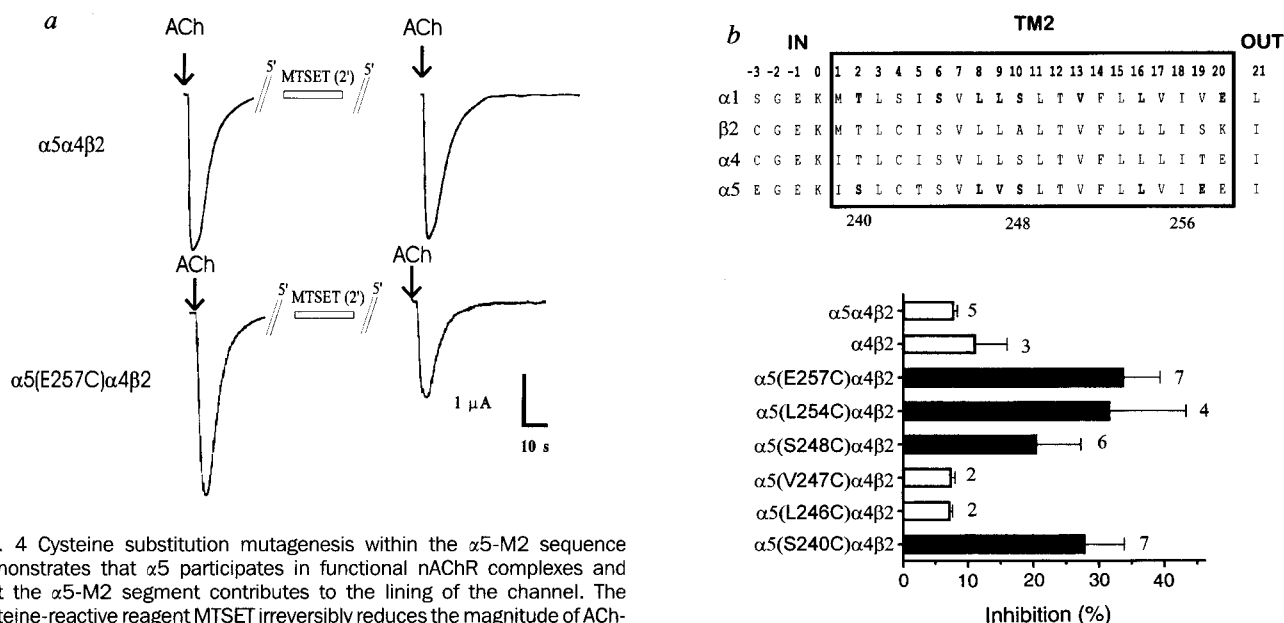


FIG. 4 Cysteine substitution mutagenesis within the $\alpha 5$ -M2 sequence demonstrates that $\alpha 5$ participates in functional nAChR complexes and that the $\alpha 5$ -M2 segment contributes to the lining of the channel. The cysteine-reactive reagent MTSET irreversibly reduces the magnitude of ACh-evoked currents in oocytes previously injected with $\alpha 5 + \alpha 4 + \beta 2$ RNA only when the $\alpha 5$ sequence has been mutated to include cysteine residues at specific sites within $\alpha 5$ -M2. **a**, Comparison of the effect of MTSET on ACh-evoked currents in oocytes coinjected with WT $\alpha 5 + \alpha 4 + \beta 2$ RNA (5 ng/0.5 ng/0.5 ng; top traces) versus those coinjected with $\alpha 5$, mutated so that a cysteine residue is substituted at $\alpha 5$ E257, $+ \alpha 4 + \beta 2$ ($\alpha 5$ E257C, 5 ng/ $\alpha 4$, 0.5 ng/ $\beta 2$, 0.5 ng; bottom traces). The effects of MTSET are tested as follows: (1) determine average peak response to 2–3 sequential applications of ACh (100 μ M; 5 min inter-test interval); (2) perfuse with control oocyte recording media (5 min); (3) perfuse with oocyte recording media with MTSET (1 mM, 2 min, indicated by open bar); (4) repeat control perfusion as in (2); (5) repeat ACh applications as in (1). The records shown in **a** are paired first or second responses to ACh before and after MTSET treatment, respectively. **b**, Summary of experiments using cysteine substitution in $\alpha 5$ -M2 to test participation of $\alpha 5$ -M2 in the channel lining. The inhibition of ACh-evoked currents by MTSET was evaluated as described in **a**. Box shows the aligned TM2 regions for mouse $\alpha 1$ and chick $\alpha 4$, $\beta 2$ and $\alpha 5$. Note that $\alpha 5$ sequence numbering is used (for example, $\alpha 5$ S240 aligns with $\alpha 1$ T244). Oocytes injected with the WT and mutant subunit cRNAs as indicated on the ordinate. Six sites within $\alpha 5$ -M2 (in bold) were mutated to a cysteine, with the sites chosen to correspond to those in $\alpha 1$ (muscle type,

also in bold) previously characterized¹⁸ using this approach. These included: $\alpha 5$ (S240C), $\alpha 5$ (L246C), $\alpha 5$ (V247C), $\alpha 5$ (S248C), $\alpha 5$ (L254C) and $\alpha 5$ (E257C). Statistically significant inhibition by MTSET was seen with 4 of the 6 sites tested with the results qualitatively similar but quantitatively less profound than those previously reported for comparable mutations in the muscle AChR $\alpha 1$ sequence¹⁸. Specifically, MTSET-inhibition of $\alpha 5$ -M2 cysteine substituted in nAChRs ranged from ~21% to 34% with the following $\alpha 5$ mutations showing ($\alpha 5$ (E257C) + $\alpha 4 + \beta 2$: $33.83 \pm 5.49\%$ ($n = 7$), $\alpha 5$ (L254C) + $\alpha 4 + \beta 2$: $31.7 \pm 11.53\%$ ($n = 4$), $\alpha 5$ (S248C) + $\alpha 4 + \beta 2$: $20.58 \pm 6.67\%$ ($n = 6$), $\alpha 5$ (S240C) + $\alpha 4 + \beta 2$: $27.95 \pm 5.92\%$ ($n = 7$)) significantly greater MTSET-inhibition (filled bars, $P < 0.05$) than $\alpha 5$ (WT) + $\alpha 4 + \beta 2$: $7.82 \pm 0.52\%$ ($n = 5$) or $\alpha 4 + \beta 2$: $11.2 \pm 4.77\%$ ($n = 3$). The other two $\alpha 5$ mutants ($\alpha 5$ (L246C) + $\alpha 4 + \beta 2$: $7.3 \pm 0.3\%$ ($n = 2$) and ($\alpha 5$ (L247C) + $\alpha 4 + \beta 2$: $7.45 \pm 0.55\%$ ($n = 2$)) show no statistically significant MTSET-inhibition (open bars). Mutants were generated using the Transformer Site Directed Mutagenesis Kit (Clontech). Recording solutions and conditions as described in the legend to Fig. 1. All data are expressed as mean $\pm$ s.e.m. and significance determined by a one way ANOVA (Microcal Origin 3.5 software).

muscle-type γ and δ subunits, participates in the binding of ACh²¹. A third possibility is that the change in EC₅₀ is an indirect effect of the changed and perhaps varying subunit composition.

Ultimately one must attempt to relate the characteristics of heterologously expressed nAChRs with those of native nAChRs in the central and peripheral nervous system. In this regard, coexpression of each of the known α - and β -subunits has failed to produce channels that resemble the higher conductance (40–60 pS) nAChR subtypes expressed in sympathetic, ciliary,

habenula, interpeduncular or hippocampal neurons of rat or chick (see ref. 5 for review). The inclusion of $\alpha 5$ with other α - and β -subunits may be important in the formation of these high-conductance nAChRs in these regions, where $\alpha 5$ is expressed. Perhaps most striking with regards to the potential physiological relevance of $\alpha 5$ -containing AChRs is the potent activation of $\alpha 5$ -containing AChRs by nicotine and the differential desensitization of these channels by submicromolar concentration of this drug (Figs 1c and 3b, c). □

Received 14 November 1995; accepted 15 January 1996.

- Boulter, J. et al. *Proc. natn. Acad. Sci. U.S.A.* **84**, 7763–7767 (1987).
- Ballivet, M. et al. *Neuron* **1**, 847–852 (1988).
- Heinemann, S. et al. *Clin. Neuropharmacol.* **14**, S45–S61 (1991).
- Lester, H. A. A. *Rev. Biophys. biomolec. Struct.* **21**, 267–292 (1992).
- McGehee, D. A. & Role, L. W. A. *Rev. Physiol.* **57**, 521–546 (1995).
- Conroy, W. G., Vemalis, A. B. & Berg, D. K. *Neuron* **9**, 679–691 (1992).
- Vemalis, A. B., Conroy, W. G. & Berg, D. K. *Neuron* **10**, 451–464 (1993).
- Conroy, W. G. & Berg, D. K. *J. biol. Chem.* **270**, 4424–4431 (1995).
- Wada, E., McKinnon, D., Heinemann, S., Patrick, J. & Swanson, L. W. *Brain Res.* **526**, 45–53 (1990).
- Winkler, J., Suhr, S. T., Gage, F. H., Thal, L. J. & Fisher, L. J. *Nature* **375**, 484–487 (1995).
- Couturier, S. et al. *J. biol. Chem.* **265**, 17560–17567 (1990).
- Elgoyhen, A. B., Johnson, D. S., Boulter, J., Vetter, D. E. & Heinemann, S. *Cell* **79**, 705–715 (1994).
- Couturier, S. et al. *Neuron* **5**, 847–856 (1990).
- Gerzanich, V., Anand, R. & Lindstrom, J. J. *molec. Pharmacol.* **45**, 212–220 (1994).
- Cooper, E., Couturier, S. & Ballivet, M. *Nature* **350**, 235–240 (1991).
- Listerud, M., Brussaard, A. B., Devay, P., Colman, D. R. & Role, L. W. *Science* **254**, 1518–1521 (1991).

- Brussaard, A. B. et al. *Pflügers Arch. Eur. J. Physiol.* **429**, 27–43 (1994).
- Akabas, M. H., Kaufmann, C., Archdeacon, P. & Karlin, A. *Neuron* **13**, 919–927 (1994).
- Villarroel, A. & Sakmann, B. *Biophys. J.* **62**, 196–208 (1992).
- Cohen, B. N., Labarca, C., Davidson, N. & Lester, H. A. *J. gen. Physiol.* **100**, 373–400 (1992).
- Karlin, A. & Akabas, M. *Neuron* **15**, 1231–1244 (1995).
- Abramson, S. M., Li, Y. & Taylor, P. J. *J. biol. Chem.* **264**, 12666–12672 (1989).
- Dennis, M. et al. *Biochemistry* **27**, 2346–2357 (1988).
- Tomaselli, G. T., McLaughlin, J. T., Jurman, M. E., Hawrot, E. & Yellen, G. *Biophys. J.* **60**, 721–727 (1991).
- Chen, J., Zhang, Y., Akk, G., Sine, S. & Auerbach, A. *Biophys. J.* **69**, 849–859 (1995).
- O'Leary, M. E. & White, M. M. *J. biol. Chem.* **267**, 8360–8365 (1992).
- Yu, C., Brussaard, A. B., Yang, X., Listerud, M. & Role, L. W. *Dev. Genet.* **14**, 296–304 (1993).
- Hamill, O. P., Marty, A., Neher, E., Sakmann, B. & Sigworth, F. J. *Pflügers Arch. Eur. J. Physiol.* **331**, 85–100 (1981).

ACKNOWLEDGEMENTS. We thank J. Riley for technical assistance and P. Flood and D. McGehee for comments on drafts of this manuscript. Supported by grants from NIH (L.R. & A.K.) and Council for Tobacco Research & McKnight Foundation (L.R.).

In vivo immunosuppression by targeting a novel protease receptor

Michel A. Duchosal*, Annette L. Rothermel†||, Patricia J. McConahey‡, Frank J. Dixon‡ & Dario C. Altieri§

* Division of Haematology, Department of Internal Medicine, Centre Hospitalier Universitaire Vaudois, Lausanne 1011, Switzerland
† The RW Johnson Pharmaceutical Research Institute, Raritan, New Jersey, USA

‡ Department of Immunology, The Scripps Research Institute, La Jolla, California 92037, USA

§ The Boyer Center for Molecular Medicine, Department of Pathology, Yale University, New Haven, Connecticut 06536, USA

|| Present address: The Boyer Center for Molecular Medicine, Department of Pathology, Yale University, New Haven, Connecticut 06536, USA

MEMBRANE receptors for blood proteases govern the clotting¹ and fibrinolytic² cascades, regulate signal transduction^{3,4}, and control the growth of mesenchymal cells^{5,6}. Despite their importance in the development of vascular injury⁷, it is unclear whether these mechanisms participate in the generation of an immune response. Here we report that targeting a factor Xa receptor, designated effector cell protease receptor-1 (EPR-1), with antisense oligonucleotide or with a monoclonal antibody (mAb 2E1)⁸ inhibited CD3/T-cell receptor-dependent lymphocyte proliferation. Immunosuppression was mediated by abolishing cytokine production and down-modulating membrane expression of the interleukin (IL)-2 receptor. *In vivo* administration of mAb 2E1 to severe-combined-immunodeficient mice injected with human peripheral blood leukocytes suppressed production of human immunoglobulin, abolished graft-versus-host disease, and protected these xenochimaeric mice from Epstein-Barr-virus-induced human lymphoproliferative disease. These observations indicate a new role for protease receptors in the regulation of the immune response, and identify a potential target for therapeutic immunosuppression in humans.

EPR-1 is an activation-dependent molecule of relative molecular mass 62,000 (*M*, 62K) which participates in protease-dependent mechanisms of lymphocyte co-stimulation^{8,9}. Anti-EPR-1 mAb 2E1 (ref. 8) nearly completely inhibited lymphocyte proliferation *in vitro* that was stimulated by soluble or plastic-immobilized anti-CD3 mitogenic mAb OKT3, by polyclonal or superantigen activators, or by unidirectional mixed lymphocyte culture (Fig. 1a, legend). This was paralleled by the ability of mAb 2E1 to abolish the release of cytokines IL-1, IL-2 and tumour-necrosis factor (TNF)- β (Fig. 1b), and to down-modulate membrane expression of the p55 subunit of the IL-2 receptor in anti-CD3-stimulated cells (not shown). With respect to known lymphocyte activation signals, mAb 2E1-mediated immuno-

suppression was entirely reversed by phorbol ester activation of protein kinase C, but exogenously added IL-1, IL-2, or engagement of the main co-stimulatory molecule CD28, had no effect (Fig. 1c). Alone or with phorbol ester, mAb 2E1 was not cytotoxic nor stimulatory for mononuclear cells. Specificity for the role of EPR-1 in this immunoregulatory pathway was substantiated by the ability of an antisense oligonucleotide targeted at the EPR-1 translational initiation sequence to inhibit CD3-dependent mononuclear cell proliferation in a dose-dependent manner (Fig. 1d). Although mAb 2E1 inhibited unfractionated mononuclear cell proliferation, it failed to diminish purified T-cell proliferation stimulated by immobilized mAb OKT3 (Fig. 1e). This suggests that accessory cells are required in this immunosuppressive pathway, and could not be substituted by extensive receptor cross-linking (Fig. 1f). The mAb 2E1 epitope on EPR-1 is spatially distinct from that recognized by co-stimulatory anti-EPR-1 mAbs⁹, further demonstrating that perturbation of different epitopes on leukocyte surface receptors may elicit opposite effects on T-cell activation¹⁰⁻¹².

To test the *in vivo* immunosuppressive effect of mAb 2E1, we used a xenochimaeric model consisting of intraperitoneal injection of human peripheral blood leukocytes (PBL) to severe-combined-immunodeficient (SCID) mice¹³⁻¹⁶. Such hu-PBL-SCID mice have high levels of human immunoglobulins in their sera, and may suffer from a human graft-versus-mouse host (GVH) disease^{16,17}. This latter complication is commonly observed when PBL are transferred according to the method currently used. When PBL originate from a donor with previous contact with the Epstein-Barr virus (EBV⁺ donor), hu-PBL-SCID mice may develop lymphoproliferative disease of human B-cell origin¹³.

SCID mice (160) injected with 50 million PBL each were analysed for serum IgG concentrations (Fig. 2, legend). All of the hu-PBL-SCID mice treated with medium alone, normal mouse serum, or anti-EPR-1 mAb 9D4, which does not diminish T-cell proliferation⁹, had mean serum levels of human IgG of 2,268 $\mu\text{g ml}^{-1}$, as measured one month after PBL transfer, with variability between groups of mice populated with PBL from each of the five donors (Fig. 2A, B). In contrast, only 1.9 $\mu\text{g ml}^{-1}$ of human IgG was found in donor-matched hu-PBL-SCID mice treated with mAb 2E1 (Fig. 2A, B). This early suppression of IgG production does not appear to result from human cell depletion caused by mAb 2E1, for three reasons. First, SCID mice given human PBL incubated *in vitro* with mAb 2E1, and injected *in vivo* with mAb 2E1, had human CD4⁺, and CD8⁺T (CD3⁺) cells, B cells, and macrophages in their mesenterium, as did donor-matched hu-PBL-SCID mice treated with mAb 9D4 (Fig. 2C). Second, a similar number of viable mononuclear cells were recovered in PBL from SCID mice 36 hours after being injected with human PBL and treated with mAb 2E1, or from donor-matched hu-PBL-SCID mice treated with mAb 9D4 (Fig. 2, legend). Fluorescence-activated cell sorting analysis on PBL from mAb 2E1-treated hu-PBL-SCID mice (group 2E1, Fig. 2, legend) revealed human (CD45⁺) leukocytes, comprising B

TABLE 1 Grading of cellular infiltrates and perivascular fibrosis characteristic of GVH in SCID mice given EBV⁺ PBL

Group	n	Lung		Kidney		Liver		Spleen	
		cell.	fibros.	cell.	fibros.	cell.	fibros.	mega.	fibros.
medium	4	1.5	1	1.25	1.5	1.25	2	2	2
2E1	6	0	0	0	0	0	0	0.7	0.5
2E1 <i>in vitro</i>	5	0.4	0.4	0.6	0.6	0.8	0.8	1	1.4
9D4	6	1.7	1	1.2	1.5	1	1.8	2	2
9D4 <i>in vitro</i>	5	2	1	0.8	1.8	1.6	2	2	2

PBL are from donor no. 5, and group labels are according to Fig. 2, legend. Level of (0, no; 1, moderate; 2, large) cellular infiltrates (cell.); (0, no; 1, moderate; 2, severe) fibrosis (fibros.), and number of megacaryocytes (mega., 0, normal; 1, severely diminished; 2, absent) are quantified according to a classification previously reported in detail¹⁶. The numbers indicate the means of grading from four to six mice (*n*) analysed per group.

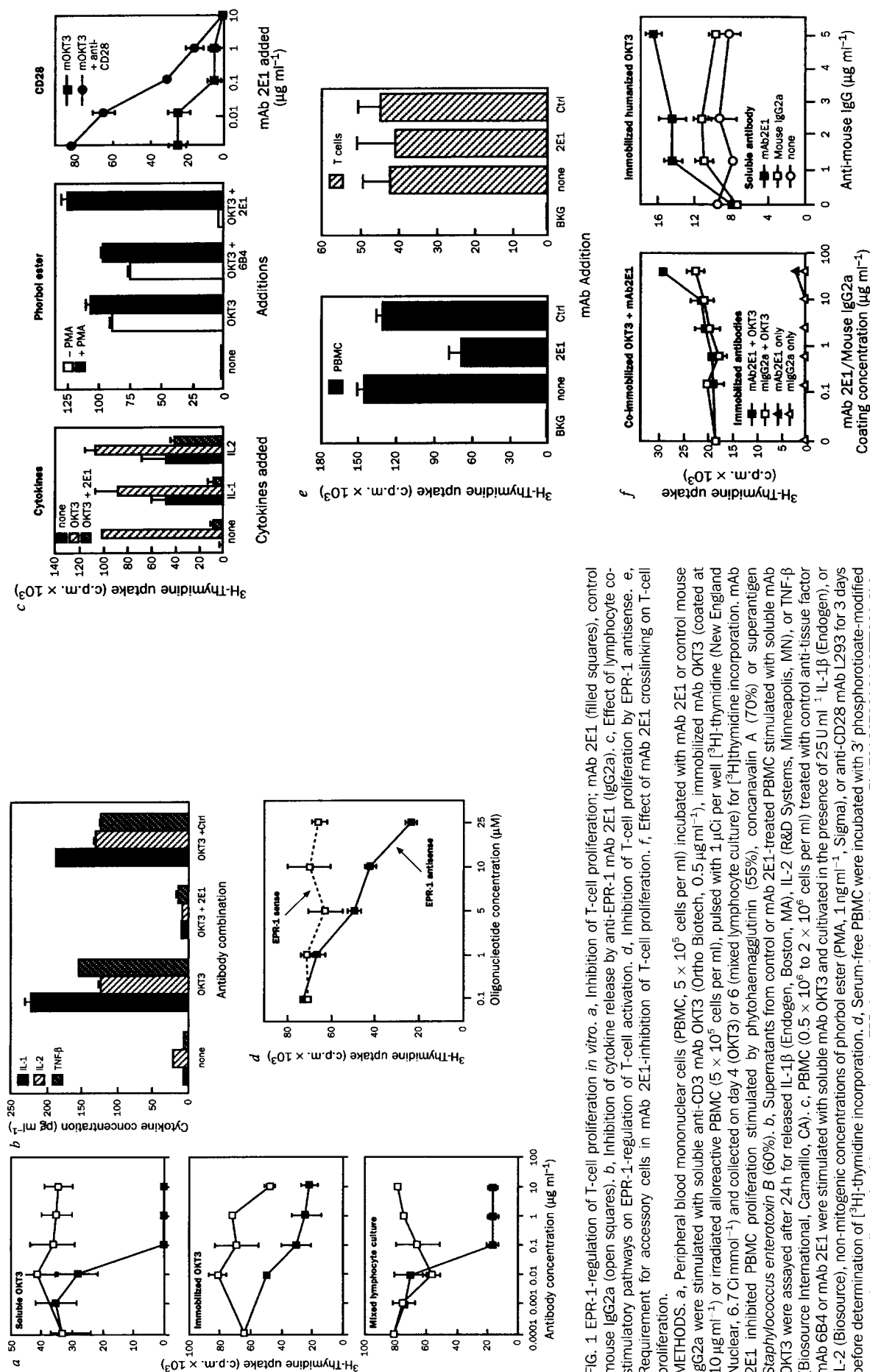


FIG. 1 EPR-1-regulation of T-cell proliferation *in vitro*. **a**, Inhibition of T-cell proliferation; mAb 2E1 (filled squares), control mouse IgG2a (open squares). **b**, Inhibition of cytokine release by anti-EPR-1 mAb 2E1 (IgG2a). **c**, Effect of lymphocyte co-stimulatory pathways on EPR-1-regulation of T-cell activation. **d**, Inhibition of T-cell proliferation by EPR-1 antisense. **e**, Requirement for accessory cells in mAb 2E1-inhibition of T-cell proliferation. **f**, Effect of mAb 2E1 crosslinking on T-cell proliferation.

METHODS. **a**, Peripheral blood mononuclear cells (PBMC, 5×10^5 cells per ml) incubated with mAb 2E1 or control mouse IgG2a were stimulated with soluble anti-CD3 mAb OKT3 (Ortho Biotech, $0.5 \mu\text{g ml}^{-1}$), immobilized mAb OKT3 (coated at $10 \mu\text{g ml}^{-1}$) or irradiated alloreactive PBMC (5×10^5 cells per ml), pulsed with $1 \mu\text{Ci}$ per well [^3H]-thymidine (New England Nuclear, 6.7 Ci mmol^{-1}) and collected on day 4 (OKT3) or 6 (mixed lymphocyte culture) for [^3H]-thymidine incorporation. mAb 2E1 inhibited PBMC proliferation stimulated by phytohemagglutinin (55%), concanavalin A (70%) or superantigen *Staphylococcus enterotoxin B* (60%). **b**, Supernatants from control or mAb 2E1-treated PBMC stimulated with soluble mAb OKT3 were assayed after 24 h for released IL-1 β (Endogen, Boston, MA), IL-2 (R&D Systems, Minneapolis, MN), or TNF- β (Biosource International, Camarillo, CA). **c**, PBMC (0.5×10^6 to 2×10^6 cells per ml) treated with control anti-tissue factor mAb 6B4 or mAb 2E1 were stimulated with soluble mAb OKT3 and cultivated in the presence of 25 U ml^{-1} IL-1 β (Endogen), or IL-2 (Biosource), non-mitogenic concentrations of phorbol ester (PMA, 1 ng ml^{-1} , Sigma), or anti-CD28 mAb L293 for 3 days before determination of [^3H]-thymidine incorporation. **d**, Serum-free PBMC were incubated with 3' phosphorothioate-modified sense or antisense oligonucleotides targeted at the EPR-1 translational initiation sequence 5'-ATGACCTCCAGAGGTTTCCCA-3' for 2 h at 37°C , stimulated with soluble mAb OKT3 and collected on day 3 for [^3H]-thymidine incorporation. **e**, PBMC or purified T cells ($> 95\%$ CD3 $^+$) incubated with control or mAb 2E1 were stimulated with immobilized mAb OKT3 (coated at $40 \mu\text{g ml}^{-1}$) and collected on day 4 for [^3H]-thymidine incorporation. **f**, T cells (5×10^5 cells per ml) were cultivated with immobilized control or mAb 2E1 with or without co-immobilized mAb OKT3, or stimulated with immobilized humanized mAb OKT3 in the presence of soluble control or mAb 2E1 plus increasing concentrations of crosslinking goat anti-mouse IgG, before measurement of [^3H]-thymidine incorporation. For all panels, data are the mean $\pm$ s.d. of a representative experiment. Each experiment was repeated at least twice with similar results.

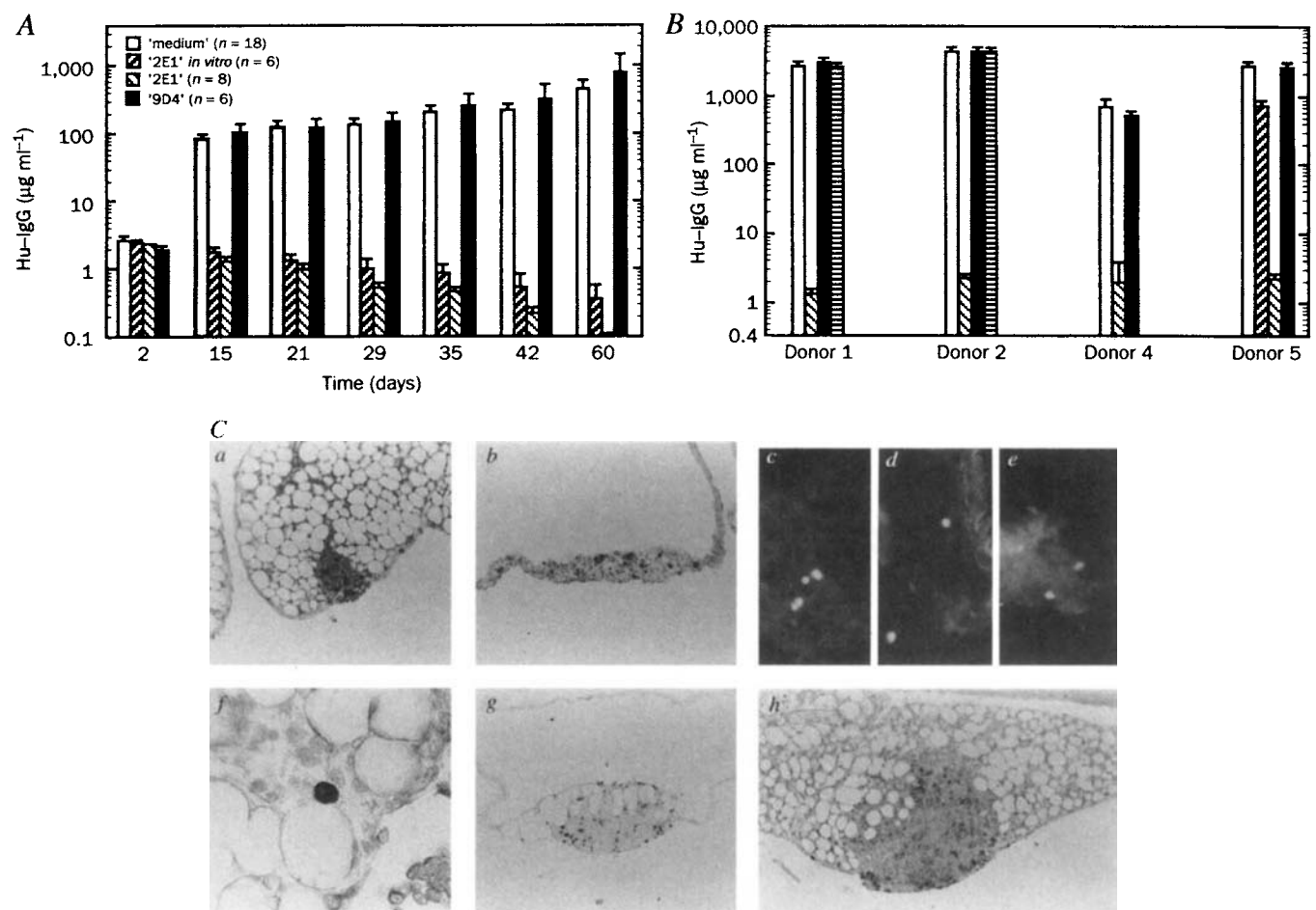
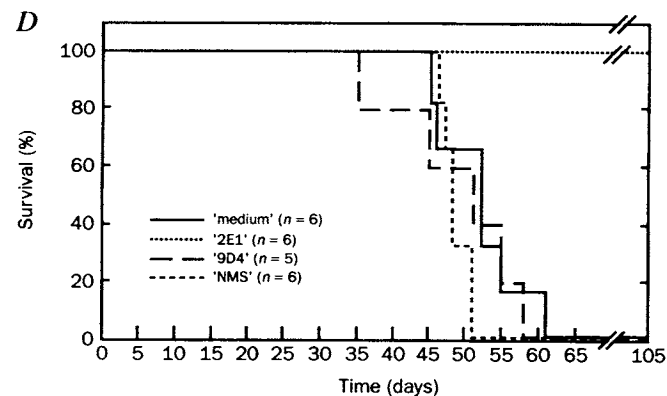


FIG. 2 Inhibition of human IgG production, detection of human cells, and survival in hu-PBL-SCID mice treated with mAb 2E1. **A**, **B**, Histograms represent mean levels ($\pm$ s.e.m.) of human IgG in available sera from: **A**, 38 SCID mice given PBL from donor no. 3 at various times after cell transfer; and **B**, 108 mice given PBL from donors no. 1, 2, 4 and 5 for one month (symbols as in **A**; horizontal bars, groups NMS). **C**, Human cells (CD45, LCA (a, b), magnification $\times 32$) comprising T cells (CD3 (c), CD4 (d), CD8 (e), $\times 64$), B cells (IgG (f) $\times 128$), and macrophages (CD68 (g), $\times 32$) in the peritoneum of SCID mice 36 h after receiving human PBL previously incubated *in vitro* with mAb 2E1, and injected *in vivo* with mAb 2E1. **h**, Human cells (9D4-LCA, $\times 32$) in a mouse given cells from the same donor, treated similarly, but with mAb 9D4. **D**, Survival curves are of 23 SCID mice populated with PBL from an EBV⁺ donor (no. 2). One mouse in group 9D4 died 2 days after cell transfer, and is not counted here. All the mice that died within 75 days after xenotransplantation had gross and microscopic evidence of lymphoproliferative disease. Hu-PBL(donor no. 2)-SCID mice treated *in vivo* with mAb 2E1 were still healthy 217 days after cell transfer, with the exception of one mouse which died of other causes. Similarly, none of the SCID mice given EBV⁺ PBL from donors no. 1, 3 and 4 and treated with mAb 2E1 had lymphoproliferative disease. The 50% survival of mice in control groups (cumulated groups medium, 9D4, and NMS) was at 93, 148, and 63 days post-PBL transfer when mice were derived from donors no. 1, 3 and 4, respectively.

METHODS. Four donors (nos. 1–4) who were Epstein–Barr seropositive [EBV⁺], and one donor (no. 5) who was Epstein–Barr seronegative [EBV[−]] gave PBL through lymphophereses after the nature and possible consequences of the study had been explained to them and they had given their informed consent. PBL were incubated in RPMI medium supplemented with 10% fetal calf serum (group medium), in same medium supplemented with $10 \mu\text{g ml}^{-1}$ of mAb 2E1 (group 2E1), of mAb 9D4 (IgG1, group 9D4), or of normal mouse serum (corresponding to $10 \mu\text{g ml}^{-1}$ of mouse IgG, group NMS). SCID mice were injected with 50×10^6 PBL each as described previously^{16,27}. Hu-PBL-SCID mice were injected intraperitoneally with Earle's medium (GibcoBRL, group medium), with $100 \mu\text{g}$ of mAb 2E1 (group 2E1), with $100 \mu\text{g}$ of mAb 9D4 (group 9D4), or with NMS (corresponding to $100 \mu\text{g}$ of mouse IgG, group NMS) weekly starting the day of cell



transfer. A minority of mice receiving cells incubated with mAb 2E1 (12 of 62), or mAb 9D4 (6 of 40) were not injected with mAb 2E1, or with mAb 9D4 *in vivo*, respectively (group 2E1 *in vitro*, and group 9D4 *in vitro*). Animal care was in accordance with institutional guidelines. In total, 24, 24, 42, 40 and 30 SCID mice received 50×10^6 PBL each from donors 1, 2, 3, 4 and 5, respectively. PBL from donors 1, 2, 3, 4 and 5 were used to generate 6, 6, 18, 10, 6 hu-PBL-SCID mice in groups medium; 6, 6, 18, 20, 12 mice in groups 2E1; 6, 6, 6, 10, 12 mice in groups 9D4; and 6, 6, 0, 0, 0 mice in groups NMS, respectively. IgG levels were determined using a method described previously²⁸. Immunohistochemistry analyses were performed using antibodies (Becton-Dickinson) to human CD3 (Leu-4), CD4 (Leu-3a), CD8 (Leu-2a) conjugated to fluorescein isothiocyanate, or to human leukocyte common antigen (LCA), IgG and IgM (Dako). Binding of unconjugated primary antibodies was indicated using commercially available kits (Vector) using diaminobenzidine as the chromogen. No binding was observed when unconjugated primary antibodies were substituted for normal rabbit serum, or a control mouse IgG1 antibody.

(CD20⁺) and T cells (CD3⁺). Finally, animals given human PBL from an EBV⁻ donor (no. 5; Fig. 2, legend) that were only exposed to mAb 2E1 *in vitro* had mean serum levels of human IgG of 215, 763 and 2,558 $\mu\text{g ml}^{-1}$ 15, 30 and 60 days after transfer. This demonstrates that human cells treated with mAb 2E1 at the time of transfer survive and remain functional in hu-PBL-SCID mice. Immunohistochemistry analysis performed on spleens, livers, kidneys, lungs and hearts from six SCID mice given human PBL from the same donor, and treated *in vivo* weekly with mAb 2E1 failed to detect human CD45⁺ cells in these animals 97 days after cell transfer. This suggests that the long-term survival of human cells in the model may be regulated by their state of activation, and/or by effector functions mediated by constant regions of antibodies used in the experiment.

Functional human T cells are essential in the development of EBV-driven human lymphoproliferative disease¹⁸, which is generally lethal for xenochimaeric animals within two months of cell transfer¹⁶, and is associated with the formation of large lymphoid tumours of B-cell origin^{13,19-22}. All 17 hu-PBL(EBV⁺)-SCID mice injected with anti-EPR-1 mAb 9D4, normal mouse serum, or medium alone died within 61 days (Fig. 2D). Autopsies performed on 13 of the 17 animals revealed large lymphoid masses in the abdomen that were histologically classified as high-grade lymphomas. In contrast, six donor-matched hu-PBL(EBV⁻)-SCID mice injected weekly with mAb 2E1 were alive and healthy 106 days after cell transfer (Fig. 2D).

Ten hu-PBL(EBV⁻)-SCID mice injected with mAb 9D4 ($n = 6$) or medium ($n = 4$) did not develop lymphoproliferative disease, but suffered from severe GVH disease, as defined by perivascular cellular infiltrates and fibrosis in multiple organs (Fig. 3a, b), severe reduction in spleen megacaryocytes (Table 1), and anaemia^{16,17}. This contrasted with six donor-matched SCID mice treated *in vitro* and *in vivo* with mAb 2E1, which remained healthy for at least 97 days after cell transfer (Table 1). At that time, histology of these animals revealed almost no evidence of GVH disease (Fig. 3c, d). Animals given human PBL that were only exposed to mAb 2E1 *in vitro* had significantly milder disease than that observed in mice treated with mAb 9D4 (Table 1).

Two general conclusions can be drawn from these studies. First, the use of a novel immunosuppressive antibody has uncovered a new regulatory aspect of lymphocyte activation that may act in concert with established pathways of T-cell co-stimulation^{23,24} to modulate antigen receptor-mediated responses. The precise mechanism of action of mAb 2E1 remains to be fully elucidated. The requirement for accessory cells in this immunosuppressive pathway suggests that EPR-1 may be implicated in alternative T-cell co-stimulatory pathways; this would be predicted by phenotypic analysis of CD28- or B7-deficient animals^{25,26}. Alternatively, mAb 2E1 might bridge accessory cell inhibitory ligand(s) to T cells. Both mechanisms may generate non-activatable T cells unable to cooperate in B-cell activation, proliferation and tumorigenesis¹⁸, and incapable of mounting a xenoreactive GVH

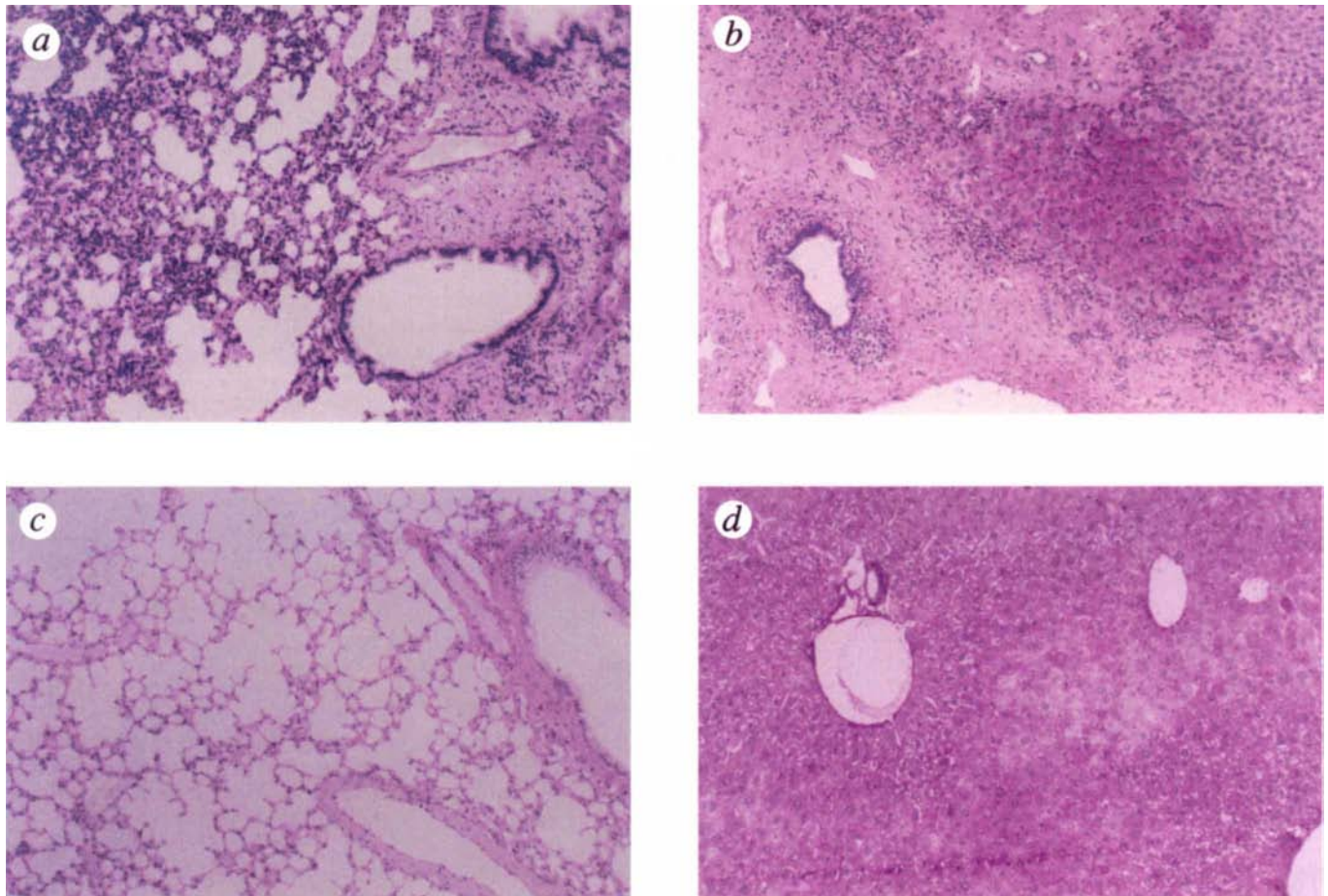


FIG. 3 Graft-versus-host disease in SCID mice given EBV⁺ PBL from donor no. 5 (Fig. 2, legend). Histological analysis of lungs (a, c) and livers (b, d) from a hu-PBL-SCID mouse in group medium 76 days after cell transfer (a, b) and a hu-PBL-SCID mouse in group 2E1 97 days after cell transfer (c, d).

METHODS. Groups of hu-PBL-SCID mice are described in Fig. 2, legend. Mouse autopsies, organ fixation, and preparation of organ sections stained with periodic acid Schiff were as described previously¹⁶. Photomicrograph: original magnification was $\times 32$.

response. In this context, repression of T-cell activation may impede long-term cell survival in the rudimentary immune environment of immunodeficient animals. Second, the immunosuppressive properties of mAb 2E1 offer a new potential alternative for target manipulation of the immune response *in vivo*. The xenochimaeric system used here is the closest known model to mimic widespread human diseases such as GVH and EBV-induced lymphomagenesis. Advanced derivatives of mAb 2E1, or synthetic antagonists of this new co-stimulatory pathway, may be beneficial for *in vivo* immunosuppression in humans. □

Received 24 November 1995; accepted 17 January 1996.

1. Davie, E. W., Fujikawa, K. & Kisiel, W. *Biochemistry* **30**, 10363–10370 (1991).
2. Pathy, L. *Cell* **41**, 657–663 (1985).
3. Vu, T.-K. H., Hung, D. T., Wheaton, V. I. & Coughlin, S. R. *Cell* **64**, 1057–1068 (1991).
4. Stitt, T. N. et al. *Cell* **80**, 661–670 (1995).
5. Glenn, K. C. & Cunningham, D. D. *Nature* **278**, 711–718 (1979).
6. Gasic, G. P., Arenas, C. P., Gasic, T. B. & Gasic, G. J. *Proc. natn. Acad. Sci. U.S.A.* **89**, 2317–2320 (1992).
7. Ross, R. *Nature* **362**, 801–809 (1993).
8. Altieri, D. C. *J. biol. Chem.* **269**, 3139–3142 (1994).
9. Altieri, D. C. & Starnes, S. J. *Cell. Immun.* **155**, 372–383 (1994).
10. Rothman, B. L. et al. *J. Immun.* **147**, 2493–2499 (1991).

11. Merkenschlager, M., Buck, D., Beverley, P. C. & Sattentau, Q. J. *Immun.* **145**, 2839–2845 (1990).
12. Walker, C., Bettens, F. & Pichler, W. J. *Eur. J. Immun.* **17**, 873–880 (1987).
13. Mosier, D. E., Gulizia, R. J., Baird, S. M. & Wilson, D. B. *Nature* **335**, 256–259 (1988).
14. Mosier, D. E., Gulizia, R. J., Baird, S. M. & Wilson, D. B. *Nature* **338**, 211 (1989).
15. Mosier, D. E., Gulizia, R. J., Torbett, B. E., Baird, S. M. & Wilson, D. B. *Nature* **353**, 509 (1991).
16. Duchosal, M. A., Eming, S. A., McConahey, P. J. & Dixon, F. J. *Am. J. Path.* **141**, 1097–1113 (1992).
17. Pirruccello, S. J. et al. *Am. J. Path.* **140**, 1187–1194 (1992).
18. Veronese, M. L. et al. *J. exp. Med.* **176**, 1763–1767 (1992).
19. Mosier, D. E. et al. *Curr. Topics Microbiol. Immun.* **152**, 195–199 (1989).
20. Okano, M. et al. *Am. J. Path.* **137**, 517–522 (1990).
21. Rowe, M. et al. *J. exp. Med.* **173**, 147–158 (1991).
22. Picchio, G. R. et al. *Cancer Res.* **52**, 2468–2477 (1992).
23. Janeway, C. A. Jr & Goldstein, P. *Curr. Opin. Immun.* **5**, 313–323 (1993).
24. Jenkins, M. K. & Johnson, J. G. *Curr. Opin. Immun.* **5**, 361–367 (1993).
25. Freeman, G. J. et al. *Science* **262**, 907–909 (1993).
26. Shahinian, A. et al. *Science* **261**, 609–612 (1993).
27. Duchosal, M. A. et al. *Nature* **355**, 258–262 (1992).
28. Duchosal, M. A., McConahey, P. J., Robinson, C. A. & Dixon, F. J. *J. exp. Med.* **172**, 985–988 (1990).

ACKNOWLEDGEMENTS. M.A.D. and A.L.R. contributed equally to this work. We thank M. Schapira for discussions; M. Rüegg and Ph. Trouillet for technical assistance; D. Xu for humanized mAb OKT3; G. Ambrosini for antisense oligonucleotides; the Blood Transfusion Center of Lausanne for blood products; the Pathology Institute (Miguel Gomez) of the Centre Hospitalier Universitaire Vaudois (CHUV) for performing tissue sections; and the staff of the Animal Facility of the CHUV for animal care. This work was supported by grants from the Swiss National Science Foundation (M.A.D.), the NIH (D.C.A.), an established investigatorship award from the American Heart Association (D.C.A.), and by Johnson & Johnson.

A histone octamer-like structure within TFIID

Alexander Hoffmann*§, Cheng-Ming Chiang*§, Thomas Oelgeschläger*, Xiaoling Xie†, Stephen, K. Burley†, Yoshihiro Nakatani‡ & Robert G. Roeder*

* Laboratory of Biochemistry and Molecular Biology, and † Laboratories of Molecular Biophysics and Howard Hughes Medical Institute, The Rockefeller University, 1230 York Avenue, New York, New York 10021, USA

‡ National Institute of Child Health and Human Development and National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, Maryland 20892, USA

THE general transcription factor TFIID nucleates initiation complex formation through direct core promoter binding^{1,2}, commits promoters within chromatin to transcription³, and mediates the action of transcriptional activators, a phenomenon that may correlate with enhanced TFIID recruitment^{4–7} or conformational changes in TFIID-promoter complexes^{8,9}. Molecular studies of the multiprotein TFIID complex have identified a primary TATA binding subunit (TBP)², TBP-associated factors (TAFs) that interact with and mediate the function of activators^{2,7,10,11} and intersubunit interactions² but have yielded relatively little insight into the structural organization of the complex or the actual mechanism of transcriptional activation. Here we present biochemical evidence for the structural relevance of histone homologies in the human TFIID subunits hTAF80, hTAF31 and hTAF20/15. Together with analyses of native TFIID complexes and accompanying crystallographic studies¹², the results suggest that there is a histone octamer-like TAF complex within TFIID.

Our database searches with the conserved domain of recently cloned human TFIID subunits TAF20/15 (refs 13; A.H. and R.G.R., manuscript submitted) and their *Drosophila* homologues dTAF28/22 (refs 14, 15) uncovered a weak sequence similarity with histone H2B (22% identity, 56% similarity) and archae-

bacterial histone-like proteins (Fig. 1). This is particularly interesting in light of previously published sequence similarities between other TFIID subunits and histones H3 and H4 (refs 15,16). Alignments incorporating predictions of secondary structure emphasize the extent of the sequence relationships over the entire histone fold (Fig. 1). Comparisons of sequence identities indicate that histone-like TAF sequences have diverged from histones as long ago as the split between archaeobacteria and eukaryotes, but after the presumed gene-duplication event that resulted in the histone H3 and H4 genes (data not shown).

To investigate the significance of proposed histone homologies we developed an assay that could reveal structural conservation of protein-protein interaction surfaces and that may be useful in examining the structural relevance of alleged histone relatedness of a growing number of proteins¹⁷. Individual histone proteins, expressed as glutathione *S*-transferase (GST) fusion proteins, were immobilized at equal concentrations (Fig. 2a) and used as substrates in binding assays with recombinant histones and corresponding histone-related TAFs. Each epitope-tagged histone displayed interactions (Fig. 2b–e) consistent with those revealed by previous biochemical and crystallographic studies with native histone complexes¹⁸. Thus H2A shows high-affinity binding to H2B (Fig. 2b, lane 4), and lower-affinity interactions with H3 and H4 (Fig. 2b, lanes 5, 6) that may contribute to the association of the H2A–H2B heterodimer with the H3–H4 tetramer. Reciprocally, H2B interacts strongly with immobilized H2A (Fig. 2c, lane 3) and more weakly with H4 (lane 6). Binding patterns observed *in vitro* with immobilized H3 (Fig. 2d) and immobilized H4 (Fig. 2e) are also similar to those observed in the native histone octamer. Although extensive biochemical studies of histones have mapped intra-octamer interactions by limited crosslinking of native histones¹⁹, we believe that our data provide the first demonstration that these protein-protein interactions can be recapitulated with recombinant proteins. This would enable mutagenesis studies of interaction surfaces to be performed in the future.

Next, full-length epitope-tagged TAF proteins produced in bacteria (f:TAF20) or baculovirus/Sf9 cells (f:TAF31 and f:TAF80) were tested for their ability to interact with immobilized histones. TAF20 interacts strongly with histone H2A and more weakly with H4 (Fig. 2f), indicating that some of its protein-interaction surfaces are similar to those of histone H2B. Similarly, f:TAF31 (Fig. 2g) and f:TAF80 (Fig. 2h) display protein-interaction potentials reminiscent of histones H3 and H4, respectively. Similar assays with truncated human TAFs constructed on the

§ Present addresses: Department of Biology, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA (A.H.); Department of Biochemistry, University of Illinois at Urbana-Champaign, 600 South Mathews Avenue, Urbana, Illinois 61801, USA (C.-M.C.).

basis of sequence alignments (Fig. 1) confirmed that histone-like interaction potentials of TAF20/15, TAF31 and TAF80 are due to domains with proposed homologies to H2B (Fig. 2i), H3 (Fig. 2j) and H4 (Fig. 2k), respectively. These observations prompted us to initiate their biophysical characterization.

Finally, we used histone-like domains in TAFs as affinity matrices (Fig. 2l) to demonstrate their capacity for histone octamer-like associations with TAF20 (Fig. 2m), TAF31 (Fig. 2n) and TAF80 (Fig. 2o). Although no H2A-homologous TAF has been identified, we found a strong TAF20 self-association (Fig. 2m, lane 3) that is detected as dimer formation in gel-exclusion chromatography of recombinant TAF20/15 (data not shown). Together with crystallographic studies that have indicated an H3/H4-like heterotetramer of the *Drosophila* homologues of hTAF80/31 (ref. 12), the results presented here suggest the potential formation of a histone octamer-like TAF complex that may be described as (hTAF20/15)₂-(hTAF80-HTAF31)₂-(hTAF20/15)₂.

To assess the physiological significance of histone-like domains in TAFs, HeLa-derived cell lines were established that stably express the epitope-tagged H2B homology domain of TAF20/15, which, *in vitro*, mediates all observed TAF20/15 protein interactions (A.H. and R.G.R., manuscript submitted). TFIID fractions from such a cell line (hf:C109, which expresses epitope-tagged TAF20(52-161)), as well as from an epitope-tagged TBP-expressing cell line (f:TBP)²⁰ and the parental control HeLa cell line, were immunoprecipitated with anti-epitope antibody. Western blotting confirmed the presence of corresponding epitope-containing polypeptides in the hf:C109 and f:TBP cell line-derived immunoprecipitates but not in that from control HeLa cells (Fig. 3a). A similar western blot probed with antigen-purified antibodies directed against TBP and a small carboxy-terminal piece of TAF20/15 demonstrates that TBP (upper open

arrow) and significant amounts of apparently full-length TAF20 and TAF15 (lower open arrows) are present, in addition to HA-FLAG-C109 (solid arrow), in the hf:C109 immunoprecipitate (Fig. 3b). A mixture of seven antisera specific to human TAFs was used in a third western blot to confirm that the hf:C109-derived complex is indeed a TFIID complex (Fig. 3c). Finally, when used in an *in vitro* transcription assay reconstituted with recombinant and highly purified general transcription factors, hf:C109-derived TFIID proved as efficient as f:TBP-IID in supporting both accurate basal and activated transcription (Fig. 3d).

These data demonstrate that the histone homology region in TAF20/15 is sufficient for incorporation into a functional TFIID complex, and that this domain of the protein is responsible for functionally important protein-protein interactions. Furthermore, that natural endogenous TAF20 and TAF15 polypeptides are co-precipitated with the tagged exogenous form indicates that there are multiple copies of TAF20/15 in the TFIID complex. Given a ~20-fold lower expression level of HA-FLAG-C109 in the hf:C109 cell line relative to endogenous TAF20/15 (data not shown), we can assume that epitope-precipitated complexes contain a single copy of the exogenous polypeptide. Phosphorimager quantification of an ¹²⁵I-protein A-labelled western blot with antibodies against shared epitopes (as in Fig. 3b) revealed that the HA-FLAG-C109 polypeptide constitutes 22% of the total TAF20/15-related polypeptides in TFIID. This is in accordance with a four-fold stoichiometry predicted by our model of a histone octamer-like substructure within TFIID.

On the basis of present results and available protein-interaction data^{13,14,16,21,22} (A.H. and R.G.R., manuscript submitted), we favour a model (Fig. 4) in which TFIID functions as a scaffold for core promoter DNA sequences that, with TFIID, constitute a docking platform for recruitment of remaining initiation factors. This model suggests some provocative hypotheses for TFIID

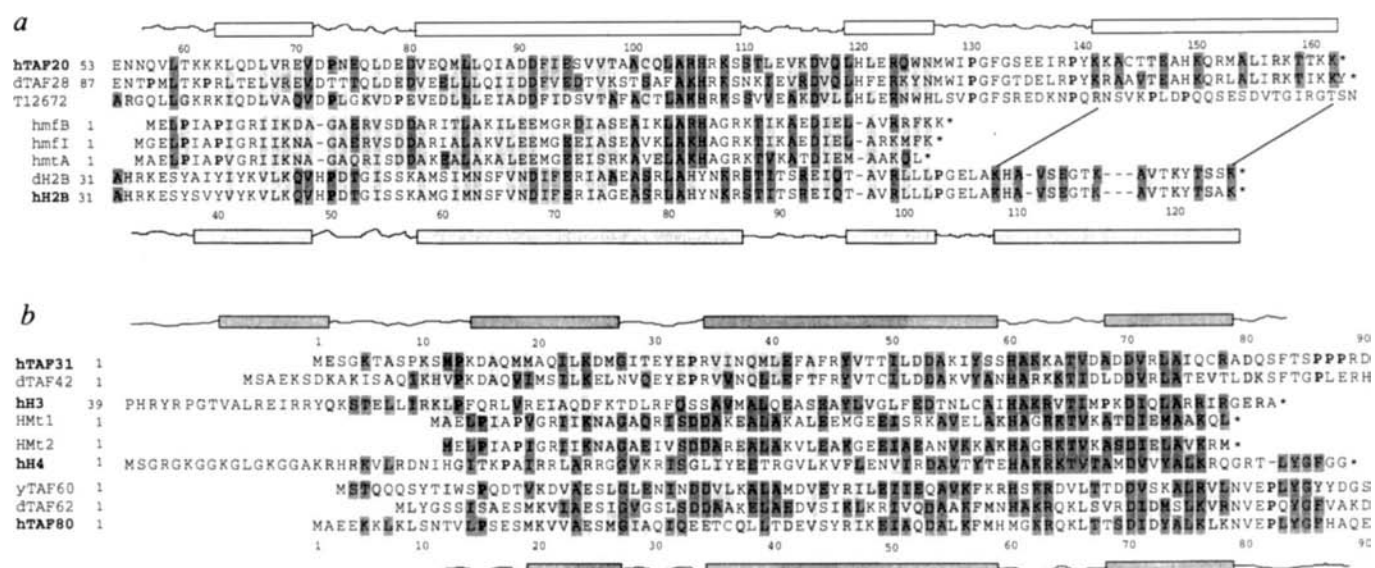


FIG. 1 Sequence similarities between TFIID subunits and histones. *a*, Sequence alignment between TAF20/15 and H2B. Alignment of partial amino-acid sequences of human TAF20/15 (A.H. and R.G.R., manuscript submitted), *Drosophila* TAF28/22 (refs 14, 15), a putative maize homologue T12672 (A.H. and R.G.R., manuscript submitted), archaeobacterial histone H2B-like proteins hmfB, hmfI and hmtA, as well as *Drosophila* and human histone H2B. Residues identical between either TAFs and any of the shown histone or histone-like proteins are darkly shaded; similar residues (in the following groupings: D, E, N; E, D, Q; K, R, H; L, T; V, I, A, M; P, G; S, T; C, S; N, Q; F, Y) are lightly shaded; potential helix breakers (prolines) are in bold; an asterisk indicates the C terminus. Empty boxes above the sequence

indicate α -helices predicted by secondary structure analysis (A.H. and R.G.R., manuscript submitted) of the hTAF20/15 and dTAF28/22 sequences; below, shaded boxes indicate α -helices identified by X-ray crystallographic analysis of the nucleosome core octamer¹⁸. *b*, Sequence alignment between TAF31/TAF80 and H3/H4. Partial amino-acid sequences of human TAF31 and TAF80 (ref. 16) as well as *Drosophila* TAF42 and TAF62 (ref. 15) and yeast TAF60 (ref. 29) are aligned with those of apparently related archaeobacterial proteins and H3 and H4 to show their mutual interrelatedness. Residues are highlighted and secondary structure is indicated as in *a*.

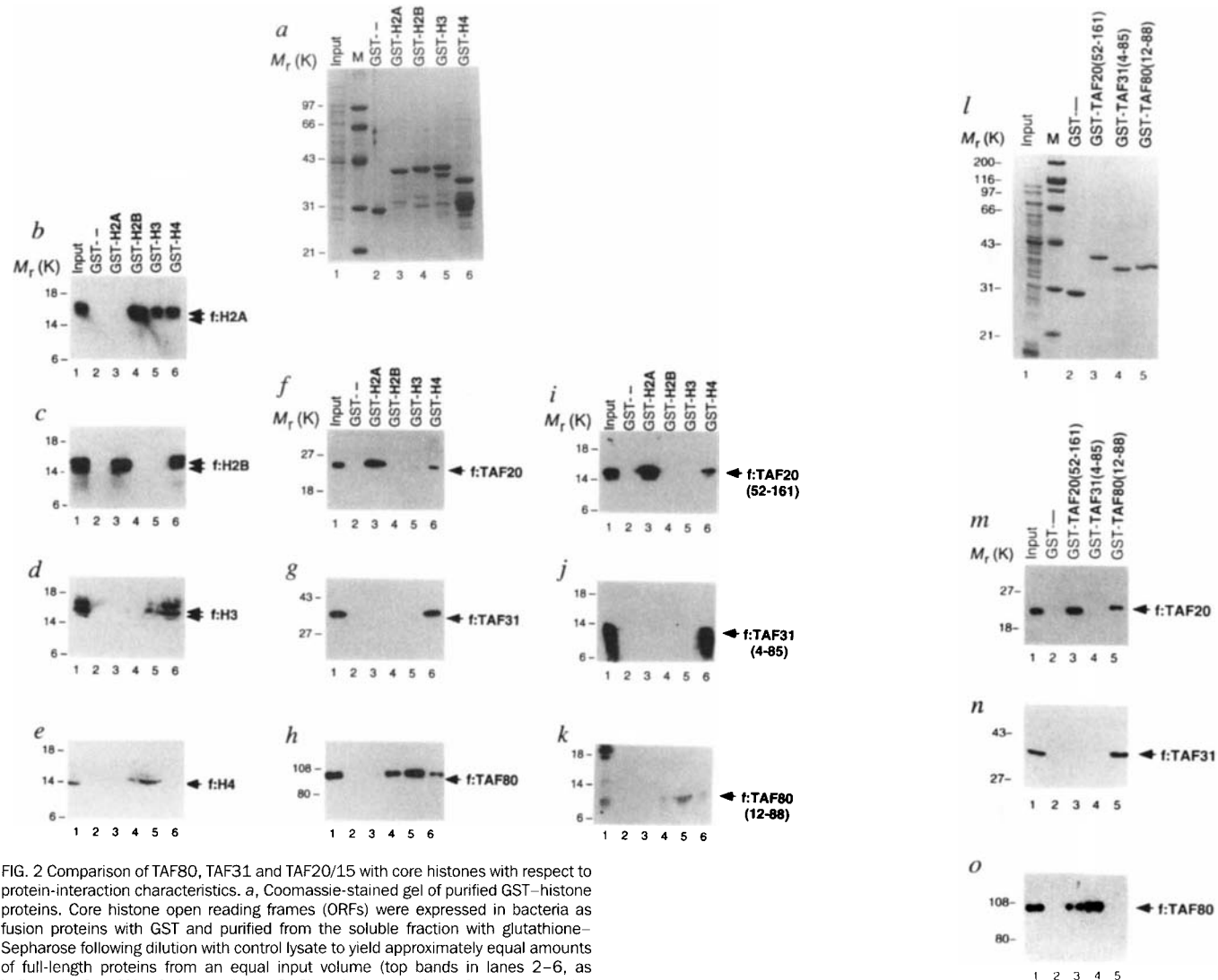


FIG. 2 Comparison of TAF80, TAF31 and TAF20/15 with core histones with respect to protein-interaction characteristics. *a*, Coomassie-stained gel of purified GST-histone proteins. Core histone open reading frames (ORFs) were expressed in bacteria as fusion proteins with GST and purified from the soluble fraction with glutathione-Sepharose following dilution with control lysate to yield approximately equal amounts of full-length proteins from an equal input volume (top bands in lanes 2–6, as indicated). One-fifth of eluates (lanes 2–6) and one-twentieth of a representative input (lane 1) are subjected to SDS-polyacrylamide gel electrophoresis (PAGE) in this and all subsequent panels. *b*–*e*, Interaction characteristics of core histones with each other. *f*–*h*, Interaction characteristics of TAF20, TAF31 and TAF80 with core histones. *i*–*k*, Interaction characteristics of the histone-like domains in TAF20/15, TAF31 and TAF80 with core histones. *l*–*o*, Coomassie-stained gel of purified GST-fusion proteins containing histone-like domains of TAF20/15, TAF31 and TAF80. Histone-like domains defined by sequence alignments (Fig. 1) were expressed in bacteria as fusion proteins with GST and purified as above. *m*–*o*, Interaction characteristics of TAF20, TAF31 and TAF80 with histone-like domains in TAFs.

METHODS. Histone ORFs were generated from genomic clones³⁰ by polymerase chain reaction (PCR). Histone homology regions of TAF80, TAF31 and TAF20/15 were

generated similarly from cDNAs¹⁶. These were inserted into pGEX-2TL(+)¹³ and FLAG-pET11d²⁰ between *Nde*I and *Bam*HI sites. The expression of histone H4 was improved by optimizing codons for the bacterial translation machinery and removing 12 CpG dinucleotides. Histone homology regions of TAF80, TAF31 and TAF20 were similarly cloned into the same expression vectors. Bacterial expression and GST-interaction assays were performed with 0.5 M NaCl with 0.2% NP40 and 0.02% sarkosyl as described elsewhere (A.H. and R.G.R., manuscript submitted) and visualized by anti-FLAG tag (M2, Kodak) western analysis. Experimental procedures were identical in all parts of this figure, and all recombinant proteins were expressed in bacteria, except f:TAF31 (*g*) and f:TAF80 (*h*), which were expressed in the baculovirus/Sf9 system.

FIG. 4 A model of the TFIID complex. This exploding view of TFIID emphasizes a nucleosome-like complex consisting of TAF80, TAF31 and TAF20/15 that is tightly linked by multiple interactions^{13,14,16,21,22} (A.H. and R.G.R., manuscript submitted) (indicated by lines) to TBP and co-activator TAFs (TAF250, TAF135, TAF55 and TAF95), some of which have been shown to contact activators. The DNA is represented by a thick line curving around the TAF-octamer. Smaller TAFs of unknown function and structure described previously¹³ are not shown here, for clarity.

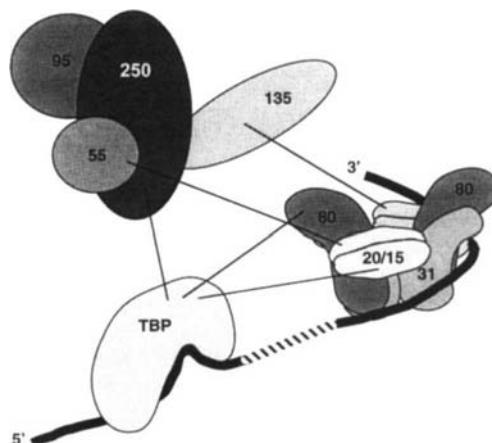
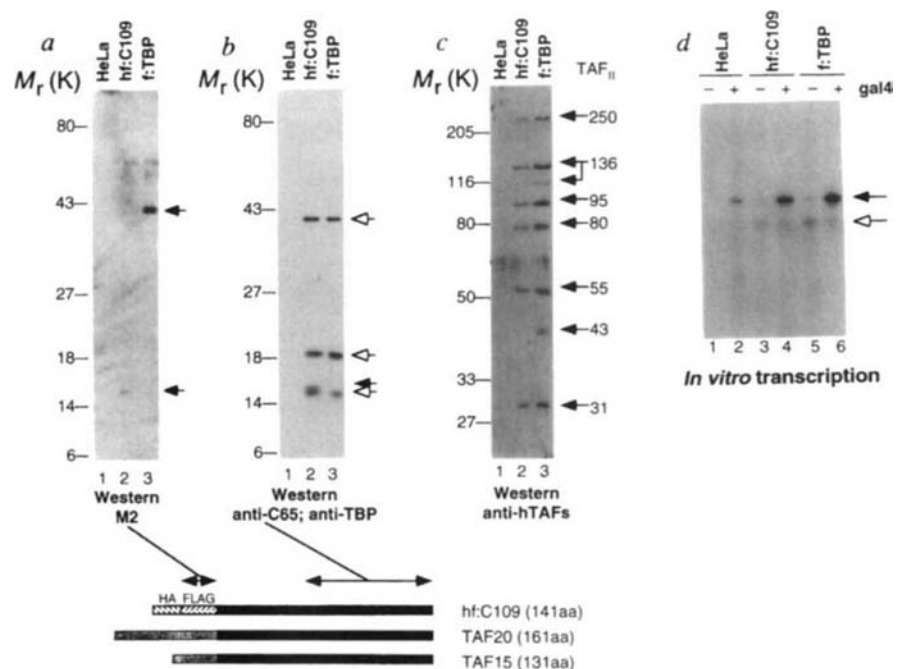


FIG. 3 Histone similarity region of TAF20/15 is sufficient for incorporation into a functional TFIID complex. **a**, Anti-FLAG western blot. Immunoprecipitates from indicated HeLa cell line-derived phosphocellulose fractions were tested for the presence of FLAG-epitope-containing proteins. The upper arrow indicates FLAG-tagged TBP (lane 3); the lower arrow indicates FLAG-tagged H2B-like domain of TAF20/15 (hf:C109, lane 2). **b**, Western blot with anti-TBP and anti-TAF20 antibodies. Immunopurified samples were tested for the presence of TBP with anti-htBP antiserum, as well as for TAF20/15 and their derivatives with antigen-purified antibodies against the C-terminal 65 amino acids of TAF20/15. Top open arrow indicates TBP, middle open arrow TAF20, and lower open arrow TAF15. Solid arrow indicates the TAF20/15 derivative hf:C109 (lane 2). **c**, Western blot with seven antisera against human TAFs. Immunopurified samples were tested for the presence of human TAFs. Antisera, each specific for a single human TAF, were first used individually and titrated (data not shown) and then mixed for the purposes of this immunoblot. Arrows point at bands of indicated TAFs. **d**, *In vitro* transcription assay. Immunopurified samples were used in an *in vitro* transcription assay to test their ability to support activator-independent (basal; lanes 1, 3 and 5) and activated transcription (lanes 2, 4 and 6). The solid arrow indicates the run-off transcript from a Gal4 site-containing promoter (pG₅HMC₂AT); the open arrow indicates core promoter (pMLA53) initiated transcripts²².

METHODS. The hf:C109 cell line was generated by retrovirus-mediated gene transfer with HA-FLAG-C109-pBABE-neo as described previously for



the f:TBP cell line 3-10 (ref. 20) also used here. Nuclear extracts were prepared, fractionated on phosphocellulose and then used for immunopurification of TFIID with M2-agarose (Kodak) as described²⁰, using tenfold more input material for each preparation of hf:C109-IID than for f:TBP-IID. *In vitro* transcription assays were performed as described²². Anti-C65 serum is described elsewhere (A.H. and R.G.R., manuscript submitted).

function. Although histone 'handshake' motifs¹⁸ within TAFs¹² may simply constitute protein-interaction domains in TFIID subunits, positive charges on the histone octamer surface lining the path of the DNA superhelix²³ are at least partly conserved in the TAFs and indicative of the possibility of TAF-DNA interactions. Indeed, early characterization of TFIID-promoter interactions likened the nuclease protection patterns of TFIID with those of nucleosomes, leading to speculations about DNA wrapping^{24,25}. In addition, a protein of relative molecular mass 60,000 (M_r 60K), potentially equivalent to dTAF62, was found to cross-link to downstream sequences of the *Drosophila* heat-shock protein 70 (hsp 70) promoter in a TATA-dependent manner²⁶.

Histone-like DNA interactions and bending may contribute to the stability of the initiation complex, and could account for the functional significance of downstream regions on weak TATA-containing²⁷ or TATA-less²⁸ promoters. Retention of TFIID components on transcriptionally inactive mitotic chromosomes (N. Segil, A.H., R.G.R. and N. Heintz, manuscript submitted) is indicative of a nucleosome-like stability of the TFIID-promoter complex within physiological chromatin. In this context, activation effects by upstream factors may be achieved through induction of conformational changes of TFIID-promoter complexes that allow more efficient formation of complete initiation complexes^{8,9}, rather than simply through TFIID recruitment *per se*⁴⁻⁷. □

Received 31 January; accepted 5 March 1996.

1. Roeder, R. G. *Trends biochem. Sci.* **16**, 402-408 (1991).
2. Burley, S. K. & Roeder, R. G. *A. Rev. Biochem.* (in the press).
3. Owen-Hughes, T. & Workman, J. L. *Crit. Rev. Euk. Gene Express.* **4**, 403-441 (1994).
4. Abmayr, S. M., Workman, J. L. & Roeder, R. G. *Genes Dev.* **2**, 542-553 (1988).
5. Workman, J. L., Abmayr, S. M., Cromlish, W. A. & Roeder, R. G. *Cell* **55**, 211-219 (1988).
6. Lieberman, P. M. & Berk, A. J. *Genes Dev.* **8**, 995-1006 (1994).
7. Sauer, F., Hansen, S. K. & Tjian, R. *Science* **270**, 1783-1788 (1995).
8. Horikoshi, M., Hai, T., Lin, Y.-S., Green, M. R. & Roeder, R. G. *Cell* **54**, 1033-1042 (1988).
9. Horikoshi, M., Carey, M. F., Kakitani, H. & Roeder, R. G. *Cell* **54**, 665-669 (1988).
10. Chen, J.-L., Attardi, L. D., Verrijzer, C. P., Yokomori, K. & Tjian, R. *Cell* **79**, 93-105 (1994).
11. Jacq, X. *et al. Cell* **79**, 107-117 (1994).
12. Xie, X. *et al. Nature* **380**, 316-322 (1996).
13. Mengus, G. *et al. EMBO J.* **14**, 1520-1531 (1995).
14. Yokomori, K., Chen, J.-L., Admon, A., Zhou, S. & Tjian, R. *Genes Dev.* **7**, 2587-2597 (1993).
15. Kokubo, T. *et al. Nature* **367**, 484-487 (1994).
16. Hisatake, K. *et al. Proc. natn. Acad. Sci. U.S.A.* **92**, 8195-8199 (1995).
17. Baxevas, A. D., Arents, G., Moudrianakis, E. N. & Landsman, D. *Nucleic Acids Res.* **23**, 2685-2691 (1995).
18. Arents, G., Burlingame, R. W., Wang, B.-C., Love, W. E. & Moudrianakis, E. N. *Proc. natn. Acad. Sci. U.S.A.* **88**, 10148-10152 (1991).
19. Pruss, D., Hayes, J. J. & Wolffe, A. P. *BioEssays* **17**, 1-10 (1995).

20. Chiang, C.-M., Ge, H., Wang, Z., Hoffmann, A. & Roeder, R. G. *EMBO J.* **12**, 2749-2762 (1993).
21. Thut, C. J., Chen, J.-L., Klemm, R. & Tjian, R. *Science* **267**, 100-104 (1995).
22. Chiang, C.-M. & Roeder, R. G. *Science* **267**, 531-536 (1995).
23. Arents, G. & Moudrianakis, E. N. *Proc. natn. Acad. Sci. U.S.A.* **90**, 10489-10493 (1993).
24. Sawadogo, M. & Roeder, R. G. *Cell* **43**, 165-175 (1985).
25. Nakajima, N., Horikoshi, M. & Roeder, R. G. *Molec. cell. Biol.* **8**, 4028-4040 (1988).
26. Gilmour, D. S., Dietz, T. J. & Elgin, S. C. R. *Molec. cell. Biol.* **10**, 4233-4238 (1990).
27. Nakatani, Y. *et al. Nature* **348**, 86-88 (1990).
28. Martinez, E. *et al. Proc. natn. Acad. Sci. U.S.A.* **92**, 11864-11868 (1995).
29. Poon, D. *et al. Proc. natn. Acad. Sci. U.S.A.* **92**, 8224-8228 (1995).
30. Zhong, R., Roeder, R. G. & Heintz, N. *Nucleic Acids Res.* **11**, 7409-7425 (1983).

ACKNOWLEDGEMENTS. We thank R. Takada and M. Horikoshi for anti-TAF250 and anti-TAF80; M. Guermah for anti-TAF31; S. Stevens for anti-TAF135 and the partial TAF135 cDNA; N. Heintz for histone genomic clones; and N. Segil, J. Darrell, A. Wolffe and J. Workman for discussions, encouragement and/or critical reading of the manuscript. This work was supported by a Boehringer-Ingelheim graduate fellowship (A.H.), by postdoctoral fellowships from the Helen Hay Whitney Foundation and the Aaron Diamond Foundation (C.-M.C.), the Deutsche Forschungsgemeinschaft (T.O.), by grants from the NIH and the Pew Trust to the Rockefeller University (R.G.R.), and the Howard Hughes Medical Institute (S.K.B.).

CORRESPONDENCE AND MATERIALS. Requests should be addressed to R.G.R. (e-mail address roeder@rockvax.rockefeller.edu).

Structural basis of calcium-induced E-cadherin rigidification and dimerization

Bhushan Nagar*, Michael Overduin†‡, Mitsuhiro Ikura†§ & James M. Rini*

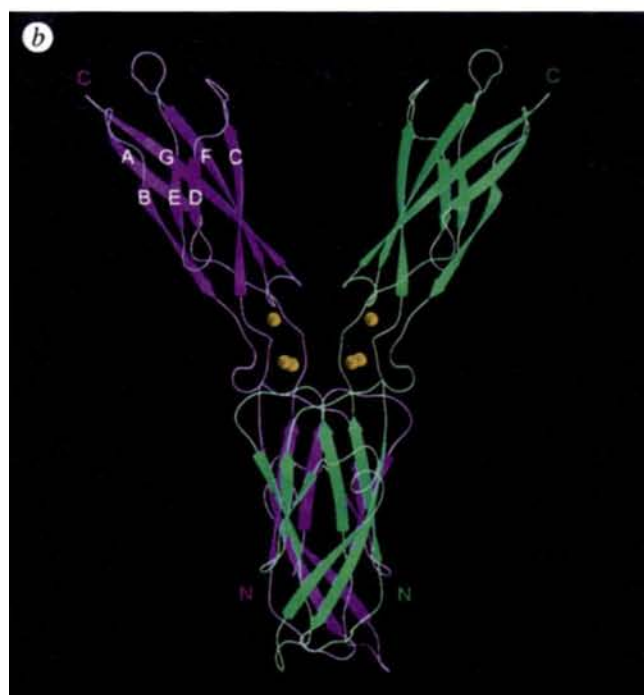
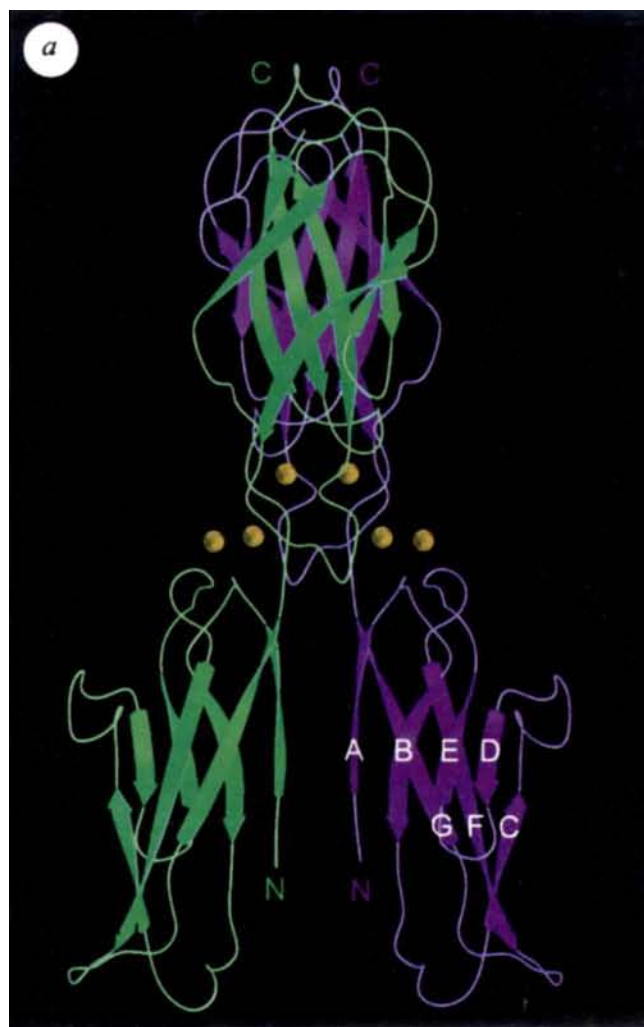
* Departments of Molecular and Medical Genetics and Biochemistry, University of Toronto, Toronto, Ontario M5S 1A8, Canada

† Division of Molecular and Structural Biology, Ontario Cancer Institute and Department of Medical Biophysics, University of Toronto, 610 University Avenue, Toronto, Ontario M5G 2M9, Canada

§ Center for Tsukuba Advanced Research Alliance and Institute of Applied Biochemistry, University of Tsukuba, Tsukuba 305, Japan

THE cadherins mediate cell adhesion and play a fundamental role in normal development¹. They participate in the maintenance of proper cell–cell contacts: for example, reduced levels of epithelial cadherin (E-cadherin) correlate with increased invasiveness in many human tumour cell types^{2,3}. The cadherins typically consist of five tandemly repeated extracellular domains, a single membrane-spanning segment and a cytoplasmic region^{4–6}. The N-terminal extracellular domains mediate cell–cell contact⁷ while the cytoplasmic region interacts with the cytoskeleton through the catenins⁸. Cadherins depend on calcium for their function: removal of calcium abolishes adhesive activity, renders cadherins vulnerable to proteases (reviewed in ref. 4) and, in E-cadherin, induces a dramatic reversible conformational change in the entire extracellular region⁹. We report here the X-ray crystal structure at 2.0 Å resolution of the two N-terminal extracellular domains of E-cadherin in the presence of calcium. The structure reveals a two-fold symmetric dimer, each molecule of which binds a contiguous array of three bridged calcium ions. Not only do the bound calcium ions linearize and rigidify the molecule, they promote dimerization. Although the N-terminal domain of each molecule in the dimer is aligned in a parallel orientation, the interactions between them differ significantly from those found in the neural cadherin (N-cadherin) N-terminal domain (NCD1) structure¹⁰. The E-cadherin dual-domain structure reported here defines the role played by calcium in the cadherin-mediated formation and maintenance of solid tissues.

A fragment of murine E-cadherin containing the two N-terminal extracellular domains (Ecad12) was expressed in *Escherichia coli* as previously described for the N-terminal domain¹¹. The protein crystallized with a calcium-mediated dimer in the asymmetric unit and the structure was solved using a combination of multiwavelength anomalous diffraction (MAD) phasing and real-space averaging techniques (Table 1). Each molecule of the dimer is composed of two 7-stranded β -barrel domains (Ecad1: residues 3–100 and Ecad2:111–213) which are not only connected by a 10-residue linker (101–110), but are also bridged by a novel arrangement of three contiguous calcium ions (Ca1, Ca2 and Ca3; Fig. 1). Ca1 and Ca3, which are associated most closely with Ecad1 and Ecad2 respectively, are linked through Ca2 by bridging interactions involving the side-chain carboxylates of Glu 11, Glu 69, Asp 103 and Asp 136 (Fig. 2a, b). The connection between Ecad1 and Ecad2 is further stabilized by the adjacent linker residues Asn 102 and Asp 103, which bind Ca3 and Ca1/Ca2 respectively. Mutation of Asp 134 to alanine inactivates E-cadherin adhesion¹² and, as seen here, would remove the bidentate interaction with Ca3. Notably, these calcium-mediated interactions account for a large number of the non-covalent interactions between Ecad1 and Ecad2. The domains do not otherwise interact very extensively, burying in total only 450 Å²



‡ Present address: Department of Pharmacology, University of Colorado Health Sciences Center, 4200 East Ninth Avenue, Denver, Colorado 80262, USA.

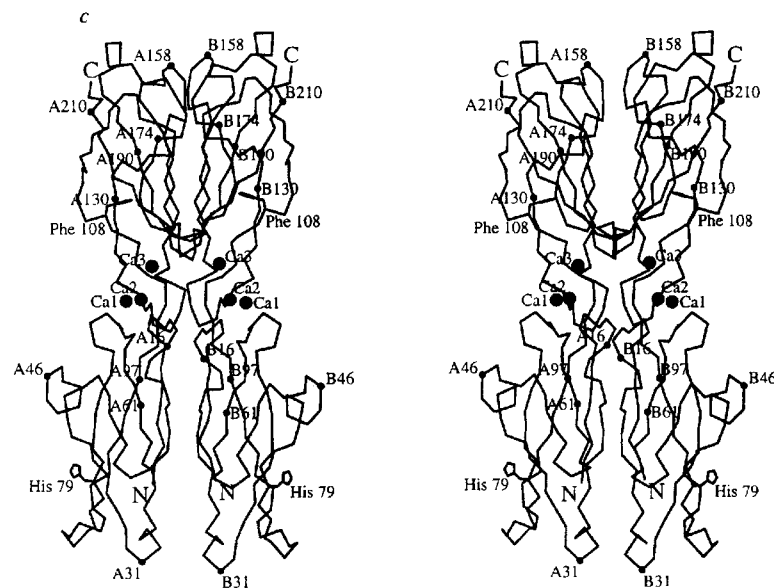


FIG. 1 Ecad12 dimer. *a*, Molecules 1 and 2 of the dimer are shown in magenta and green, respectively. The molecules of the dimer are related by a non-crystallographic 2-fold symmetry axis running vertically in the plane of the page. The cluster of three calcium ions (yellow) is bound in the linker region connecting the N- and C-terminal domains of each molecule and, in this view, are separated by the 2-fold axis. In both cases, Ca1 and Ca3 are the outermost and uppermost, respectively. The N- and C-terminal domains are composed of 7-stranded β -barrels (A–G) showing the same topology and a very similar three-dimensional structure (r.m.s. deviation of 2.0 Å for 88 corresponding C α positions). There is a short α -helical segment connecting strands E and F in both domains. In Ecad1, the connection between strands B and C (residues 27–32) is disordered. The connection between strands C and D in Ecad1 contains two consecutive proline residues and assumes a conformation similar to that termed a quasi- β helix in the NCD1 structure¹⁰. *b*, The same dimer as in *a*, rotated by $\sim 90^\circ$ around the two-fold symmetry axis. In this view, the C-termini are splayed apart and directed away from the 2-fold symmetry axis. *c*, Stereo C α plot rotated $\sim 45^\circ$ around the two-fold symmetry axis relative to that in *a*. The N-terminal residues seen in the electron density correspond to Val 3. The side chain of His 79 of the cadherin HAV adhesion motif is shown. The conformation of the HAV motif and the surrounding region is very similar to that seen in the NCD1 structure. In Ecad12 this motif is involved in a translational crystal lattice contact unlikely to be of physiological relevance. Note that both the NCD1 and the Ecad12 constructs represent only a portion of a complex membrane-spanning cell-adhesion molecule, and neither have been tested in functional assays. The side chain of Phe 108, proposed as a possible mediator of cadherin parallel dimer interactions¹⁰, is seen to form part of the hydrophobic core of Ecad2.

between them. Indeed, relatively sharp NMR linewidths indicate that, in the absence of calcium ions, the two domains of Ecad12 tumble relatively freely in solution (data not shown). The network of calcium-mediated interactions would undoubtedly fix the orientation between Ecad1 and Ecad2 when calcium is bound, thus rigidifying the molecule. This network of interactions further suggests that calcium binding would be a highly cooperative process (also see ref. 9). Given that extracellular calcium levels are regulated over a relatively small concentration range and yet serve a signalling role in many organisms¹³, cooperative calcium-binding properties may be of functional significance.

Residues in the sequence motifs PENE (10–13), LDRE (66–69), DQNDN (100–104), and DAD (134–136) have been implicated in calcium binding^{6,14,15} and we now show that Asn 143 and Asp 195 are important as well (Fig. 2*b*). Based on sequence homology and the detailed structure of the calcium-binding region, we predict that calcium is bound in a similar fashion

between each of the tandemly repeated extracellular domains, resulting in a stoichiometry of 12 calcium ions per molecule. Calcium binding would be expected to induce rigidification along the entire length of the extracellular region, a structural feature likely to be shared by all members of the cadherin superfamily. Interestingly, mutations in the PENE¹⁶ and DAD¹⁷ motifs of Ecad2/Ecad3 and Ecad4/Ecad5, respectively, have been found in two human carcinomas.

Calcium binding is required for cadherin-mediated cell adhesion, and we show here that it is also central to E-cadherin dimer formation. Through interactions with residues in the linker region, calcium binding structures a large component of the 2-fold symmetric dimer interface. Of particular importance are the direct hydrogen bonds made between molecules of the dimer by the side chains of Gln 101 and Asn 143, residues that bind calcium through their backbone carbonyl oxygen atoms (Fig. 2*c*). Conversely, the side chain of Asp 100 coordinates calcium, while its carbonyl oxygen atom mediates a direct hydrogen bond across the dimer interface. In addition, Asp 100, Gln 101, Asn 102 and Asn 143 stabilize dimer formation through water-mediated hydrogen-bond interactions. Dynamic light scattering and sedimentation

analysis show that Ecad12 dimerizes only in the presence of calcium ($K_d = 0.6$ mM in 10 mM calcium; data not shown). In this regard, Ecad12 differs from the N-cadherin NCD1 fragment, which dimerizes even in the absence of calcium¹⁰.

The N-terminal domains of the Ecad12 dimer are aligned in a parallel orientation (Fig. 1*a*), an arrangement similar to that seen in the NCD1 X-ray crystal structure¹⁰. Although this similarity suggests that parallel dimer formation is common among cadherins, in E-cadherin the N-terminal domains are far less intimately associated. There are no direct hydrogen bonds between the paired Ecad1 domains, and Trp2, which in the NCD1 structure crosses over to the adjacent molecule (described as a strand exchange), is disordered in the Ecad12 dimer. Furthermore, only 1,000 Å² of the total surface area buried (2,000 Å²) in the Ecad12 dimer interface is accounted for by residues in the N-terminal domains, whereas the equivalent residues in NCD1 bury $\sim 1,800$ Å². Although the N-terminal domains of Ecad12 and NCD1 are quite similar (0.75 Å r.m.s.d. for all C α atoms), the angle between them in their respective parallel dimers also differs (by about 56°). Consequently, in the NCD1 dimer orientation, the C termini would be too far apart (~ 24 Å) to allow residues in the linker region to interact across the dimer interface as seen here. These differences presumably arise from the fact that the NCD1 structure represents only the N-terminal domain and therefore lacks the complete calcium-binding region. The apparently fortuitous and more extensive interaction between the N-terminal domains in the more truncated NCD1 construct might account for its ability to dimerize in a calcium-independent fashion. In Ecad12, interactions across the dimer interface are mediated, more importantly, through residues structured by bound calcium ions.

Based on the assumption that the orientation between all five E-cadherin tandem repeats would be the same as that observed in Ecad12, we generated a model for the entire extracellular region. As shown in Fig. 3, the molecules of the dimer would assume a V-shaped arrangement. Each arm is ~ 240 Å in length, very close to the 220 Å obtained by electron microscopy⁹ for the extracellular region, in the presence of calcium. Although speculative, this model raises the possibility that the extracellular domains of a cadherin dimer do not interact beyond what is described here for the Ecad12 structure.

Cell–cell adhesion, mediated by the classic cadherins, involves the conserved HAV motif^{7,18} located on the F strand of the N-terminal domain (Fig 1*a*, *c*). In the NCD1 crystal lattice, the parallel dimers were found also to make antiparallel interactions,

a

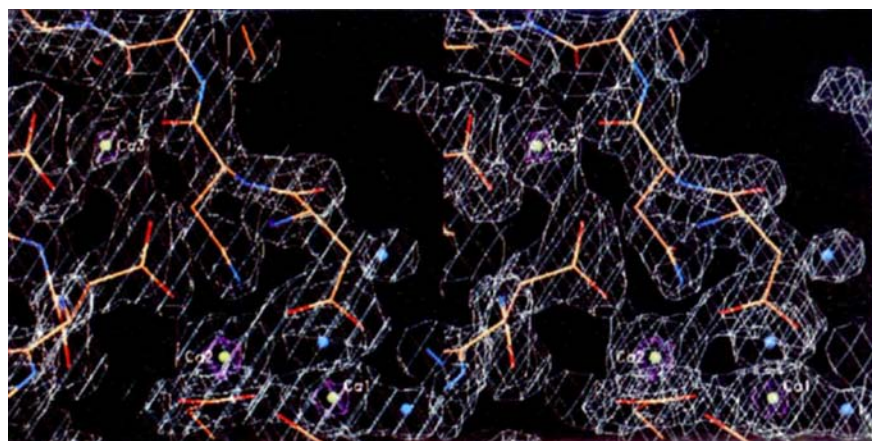


TABLE 1 Structure determination and refinement

	Nat	Hg1	Pb1	Hg2 (MAD)			Hg3
Diffraction data							
Heavy-atom reagent	–	MMA	TMLA		MMA		MMA
Wavelength (Å)	1.5419	1.5419	1.5419	1.0110	1.0095	1.0073	0.908
Resolution (Å)	4.0	4.0	4.0	2.5	2.5	2.5	2.0
Temperature (°C)	20	20	20	–160	–160	–160	–160
Measured reflections (no.)	14,672	16,368	10,023	36,104	37,522	36,320	86,202
Unique reflections (no.)	4,790	5,192	4,252	23,998	24,834	24,362	37,605
Completeness (%)	86	93	76	59	60	60	88
R_{sym} *	0.051	0.121	0.147	0.050	0.052	0.052	0.053
Number of sites	–	2	2	2	2	2	2
Phasing power†, isomorphous	–	2.1	2.2	–	1.4	1.4	–
anomalous	–	–	–	1.5	1.9	2.1	1.8
Refinement statistics							
Resolution (Å)	5.0–2.0		R.m.s.d. bond length (Å)				0.010
No. of reflections	33,174		R.m.s.d. bond angle (°)				1.591
R_{cryst}	0.204		R.m.s.d. B-values (Å ²)				2.12
R_{free}	0.267						

Crystals were grown by the vapour-diffusion method from protein (11 mg ml^{−1}) drops containing 20 mM Tris-HCl, pH 7.25, 5 mM dithiothreitol, and 50 μM sodium azide, equilibrated against wells (1 ml) containing 1.2 M ammonium sulphate, 5–10 mM calcium chloride, 100 mM Tris-HCl, pH 9.0, and 3 mM sodium azide. Plate-like crystals grew in space group C2 ($a = 122.0$ Å, $b = 80.5$ Å, $c = 73.2$ Å, $\beta = 114.5^\circ$) with a dimer in the asymmetric unit (63% solvent). Attempts to grow crystals in the absence of calcium were unsuccessful. Derivatives were obtained by soaking crystals in artificial mother liquor (1.4 M ammonium sulphate, 5–30 mM calcium chloride, 100 mM Tris-HCl, pH 9.0) containing 1 mM methyl mercury acetate (MMA) for 1 d, or 20 mM trimethyl lead acetate (TMLA) for 10 d. The native (Nat) and derivative data sets, Hg1 and Pb1, were collected on a Siemens multiwire area detector mounted on a Rigaku rotating anode at room temperature and processed with XDS²⁰. Derivative data sets Hg2 (collected at the Hg L_{III} edge) and Hg3 were collected on the Cornell High Energy Synchrotron Source beamlines F2 and A1, respectively, using a CCD area detector. Both synchrotron data sets were collected at -160°C from crystals flash-frozen in liquid propane (cryoprotected in artificial mother liquor containing 40% glycerol) and processed using DENZO²¹. Values of f' and f'' for mercury were taken to be those determined experimentally for a mercury–cysteine complex²². All phase calculations were performed using PHASES²³. As only 60% (in all resolution ranges) of the Hg2 data set had been collected (Friedel mates were obtained for each measurement using inverse beam geometry), MAD maps calculated from these data were not readily interpretable. As this data set contained the best high-resolution phase information (>4.0 Å), a procedure, to be described in detail elsewhere, was implemented to fill in the missing data using the non-isomorphous data sets Nat/Hg1/Pb1 and Hg3. In short, a least-squares procedure was developed to refine the masked 4.0 Å-resolution multiple isomorphous replacement (MIR) density (using Nat, Hg1, and Pb1), corresponding to both domains of both molecules, against the amplitudes of the Hg2 and Hg3 data sets. The rotated and translated electron density yielded intercrystal averaging matrices as well as masks for each of the four domains in the Hg2 and Hg3 cells. Using these masks and experimentally derived phase estimates (MIR, MAD and single anomalous scattering (SAS)) for each of the three crystals (Nat/Hg1/Pb1, Hg2 and Hg3), several cycles of intercrystal averaging, two-fold symmetry averaging, solvent-flattening, back-transformation and phase combination to 4.0 Å resolution were implemented. The resultant map back-transformed in the Hg2 cell generated the missing amplitudes and phases to 4.0 Å resolution. Using the modified Hg2 data set (now complete to 4.0 Å resolution, but still only 60% complete in the 4.0–3.0 Å resolution shell), a cyclical process of solvent-flattening, two-fold symmetry averaging, back-transformation and phase combination was then carried out to refine the MAD phase probability distributions to 3.0 Å resolution. A similar procedure incorporating intercrystal averaging with the Hg3 data set (thereby incorporating its amplitude and SAS phase information to 3.0 Å resolution) was also performed. In both cases, the resultant electron density maps showed an almost continuous backbone of electron density and very well defined side chains. The model was built using program O (ref. 24) and refined against data set Hg3 (88% complete after averaging Bijvoet pairs) using X-PLOR²⁵. In the final stages of refinement, constraint interactions with the calcium ions were removed. The final model contains 422 amino acids (3,228 atoms), 6 calcium ions, 2 mercury atoms (one per molecule), and 265 water molecules. The average temperature factors were 26.8 Å² and 30.2 Å² for the protein-backbone and side-chain atoms respectively, and 16.2 Å², 32.7 Å² and 30.1 Å² for the calcium ions, water molecules and mercury atoms. The occupancies of Ca1, Ca2, Ca3 and mercury, averaged over both molecules, are 0.84, 1.0, 1.0 and 0.6, respectively. Figures were generated with SETOR²⁶ and MOLSCRIPT²⁷.

* $R_{\text{sym}} = \sum |I - \langle I \rangle| / \sum I$, where I is the observed intensity, and $\langle I \rangle$ is the average intensity obtained from multiple observations of symmetry-related reflections.

† Phasing power, r.m.s. $F_H/r.m.s.(\epsilon)$, where ϵ is lack of closure and F_H is the calculated heavy-atom structure factor.

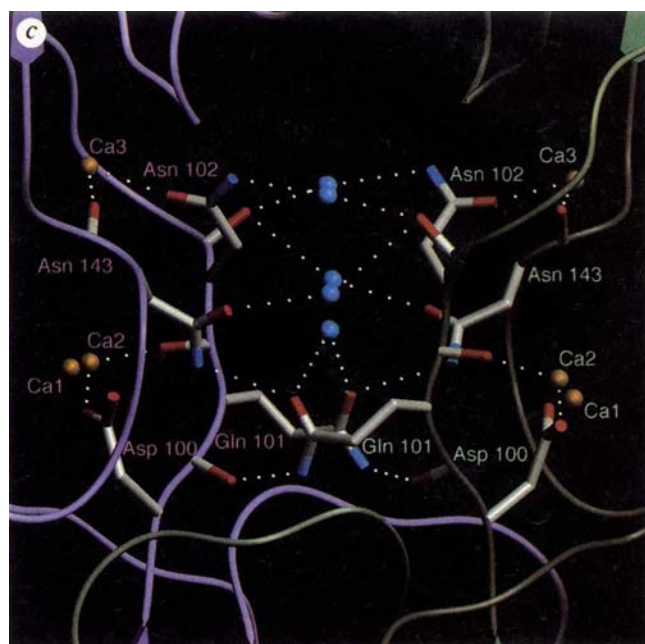
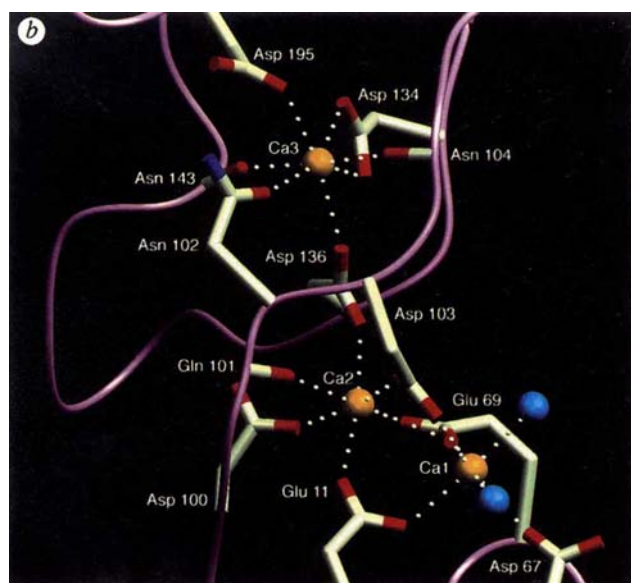


FIG. 2 The calcium-binding region. *a*, Stereo diagram of the final $2F_o - F_c$ electron-density map in the calcium-binding region of molecule 1. The map was calculated using data in the 8.0–2.0 Å resolution range and contoured at 1.8σ (grey). Also shown is the calcium-omit difference electron density contoured at 16σ (magenta). Calcium ions (Ca1, Ca2 and Ca3) and water molecules are shown as yellow and blue spheres, respectively. *b*, Calcium-binding site of molecule 1 showing the bound calcium ions and the calcium-coordinating side chains and backbone carbonyl oxygen atoms. The orientation of the molecule is the same as that of the magenta molecule in Fig. 1*a*. Ca2 and Ca3 are each coordinated by 7 oxygen atoms (pentagonal bipyramidal coordination geometry), whereas Ca1 is coordinated by six, two of which are from water molecules (blue spheres).

The coordination geometry, exclusive use of oxygen atoms, and the mean calcium–ligand bond distances of 2.29 ± 0.17 , 2.40 ± 0.17 , 2.42 ± 0.08 for Ca1, Ca2 and Ca3, respectively (averaged over both molecules), are typical of those seen in other protein–calcium complexes²⁸. *c*, Calcium-mediated dimer interactions. The orientation here is similar to that in Fig. 1*b*. Shown are selected hydrogen-bond and side-chain interactions that demonstrate the involvement of calcium in mediating dimer formation. Carbonyl oxygen atoms are shown for each residue whose side chain is depicted.

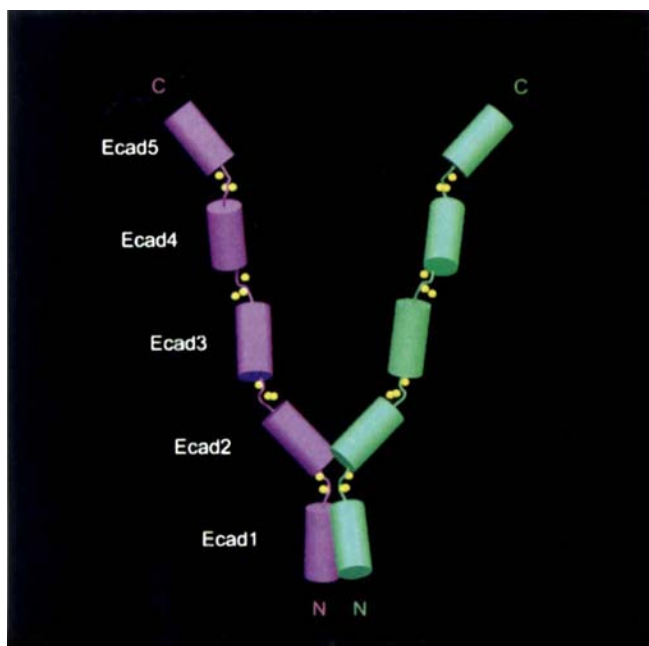


FIG. 3 The E-cadherin extracellular region. Coordinates for the extracellular region were generated by successive superimpositions of Ecad1 on Ecad2. Each domain was then rendered as a cylinder. Calcium ions are represented by oversized yellow spheres. The C termini are connected to the transmembrane segments. The lack of pairing beyond Ecad2 is a direct consequence of the fact that Ecad2 is rotated by $\sim 130^\circ$ relative to Ecad1, placing the start of the Ecad2/Ecad3 linkers ~ 64 Å apart, on the 'outside' of the dimer (Fig. 1*b*). This orientation precludes the possibility of making interactions similar to those found in the Ecad12 linker region. Consistent with the lack of analogous linker interactions is the fact that Gln 101 and Asn 143, residues whose side chains make direct hydrogen bonds across the dimer interface in Ecad12, are not conserved among the other E-cadherin extracellular domains.

involving this motif, leading to a zipper-like model for cadherin-mediated cell adhesion¹⁰. These antiparallel 'adhesive interactions' are not seen in the Ecad12 crystal structure, although sterically Ecad12 could be accommodated in the zipper-like arrangement. Regardless, we have shown that the building block of any higher-order organization is a parallel cadherin dimer whose structure is promoted by and dependent on the presence of bound calcium ions. The dimerization, rigidification and resulting mechanical stabilization of E-cadherin explains, at least in part, the calcium requirement for the integrity of cell junctions¹⁹. □

Received 27 November 1995; accepted 1 February 1996.

1. Takeichi, M. *Science* **251**, 1451–1455 (1991).
2. Mareel, M., Bracke, M. & Van Roy, F. *Molec. Biol. Rep.* **19**, 45–67 (1994).
3. Birchmeier, W. *Bioessays* **17**, 97–99 (1995).
4. Takeichi, M. A. *Rev. Biochem.* **59**, 237–252 (1990).
5. Blaschuk, O. W., Munroe, S. B. & Farookhi, R. *Can. J. Oncol.* **4**, 291–301 (1994).
6. Overduin, M. et al. *Science* **267**, 386–389 (1995).
7. Nose, A., Tsuji, K. & Takeichi, M. *Cell* **61**, 147–155 (1990).
8. Kemler, R. *Trends Genet.* **9**, 317–321 (1993).
9. Pokutta, S., Herrenknecht, K., Kemler, R. & Engel, J. *Eur. J. Biochem.* **223**, 1019–1026 (1994).
10. Shapiro, L. et al. *Nature* **374**, 327–337 (1995).

11. Tong, K. I. et al. *FEBS Lett.* **352**, 318–322 (1994).
12. Ozawa, M., Engel, J. & Kemler, R. *Cell* **63**, 1033–1038 (1990).
13. Brown, E. M., Vassilev, P. M. & Hebert, S. C. *Cell* **83**, 679–682 (1995).
14. Ringwald, M. et al. *EMBO J.* **6**, 3547–3653 (1987).
15. Overduin, M., Tong, K. I., Kay, C. M. & Ikura, M. *J. Biomol. NMR*, (in the press).
16. Oda, T. et al. *Proc. natn. Acad. Sci. U.S.A.* **91**, 1858–1862 (1994).
17. Rüscher, J. L., Berchuck, A., Kohler, M. F. & Boyd, J. J. *Nature Genet.* **7**, 98–102 (1994).
18. Blaschuk, O. W., Sullivan, R., David, S. & Pouliot, Y. *Dev Biol* **139**, 227–229 (1990).
19. Volberg, T., Geiger, B., Kartenbeck, J. & Franke, W. W. *J. Cell Biol.* **102**, 1832–1842 (1986).
20. Kabsch, W. J. *appl. Crystallogr.* **21**, 916–924 (1988).
21. Otwinowski, Z. in *Proceedings of the CCP4 Study Weekend* (eds Sawyer, L., Isaacs, N. & Bailey, S.) 56–62 (SERC Daresbury Laboratory, Warrington, UK, 1993).
22. Tesmer, J., Stemmler, T., Penner-Hahn, J., Davisson, V. & Smith, J. *Proteins* **18**, 394–403 (1994).
23. Furey, W. & Swaminathan, S. *Meth. Enzym.* (in the press).
24. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, A. *Acta crystallogr. A* **47**, 110–119 (1991).
25. Brünger, A. X-PLOR, v3.1 A System for X-ray Crystallography and NMR (Yale University Press, New Haven, USA, 1992).
26. Evans, S. V. J. *molec. Graph.* **11**, 134–138 (1993).
27. Kraulis, P. J. *appl. Crystallogr.* **24**, 946–950 (1991).
28. Glusker, J. P. *Adv. Prot. Chem.* **42**, 1–76 (1991).

ACKNOWLEDGEMENTS. We thank M. Takeichi for the E-cadherin cDNA clone and for discussion and encouragement, C. Kay for unpublished equilibrium-sedimentation and dynamic light-scattering data, and K. I. Tong and P. Yau for help in subcloning and for protein expression and purification. This work was supported by grants from the National Cancer Institute of Canada to J.M.R., M.I. and M.O. Coordinates will be deposited in the Brookhaven Protein Data Bank.

CORRESPONDENCE and requests for materials to be addressed to J.M.R. (e-mail rini@gene4d.med.utoronto.ca).

Organ targeting *in vivo* using phage display peptide libraries

Renata Pasqualini & Erkki Ruoslahti

La Jolla Cancer Research Center, The Burnham Institute, 10901 North Torrey Pines Road, La Jolla, California 92037, USA

PREFERENTIAL homing of tumour cells^{1,2} and leukocytes^{3,4} to specific organs indicates that tissues carry unique marker molecules accessible to circulating cells. Organ-selective address molecules on endothelial surfaces have been identified for lymphocyte homing to various lymphoid organs and to tissues undergoing inflammation^{5–8}, and an endothelial marker responsible for tumour homing to the lungs has also been identified⁹. Here we report a new approach to studying organ-selective targeting based on *in vivo* screening of random peptide sequences. Peptides capable of mediating selective localization of phage to brain and kidney blood vessels were identified, and showed up to 13-fold selectivity for these organs. One of the peptides displayed by the brain-localizing phage was synthesized and shown to specifically inhibit the localization of the homologous phage into the brain. When coated onto glutaraldehyde-fixed red blood cells, the peptide caused selective localization of intravenously injected cells into the brain. These peptide sequences represent the first step towards identifying selective endothelial markers, which may be useful in targeting cells, drugs and genes into selected tissues.

We injected phage libraries^{10–13} intravenously into mice and subsequently rescued the phage from individual organs. Within the time frame of the experiments, the bulk of the phage remained in circulation (not shown). Some organs, such as liver and lung, captured too many phage to be used as target organs for selection; we focused on peptide sequences that directed phage binding to the brain and kidney, because these organs bound relatively few phage from the unselected libraries.

To select peptides that home to the brain, phage were injected intravenously, recovered from the brain, amplified repeatedly *in vitro*, and re-injected to obtain sufficient enrichment. Although more of the injected phage were recovered from an equivalent amount of kidney than from brain after the first injection, 6- and 13-fold more phage were recovered from the brain in the second

and third rounds of the selection, respectively. This enrichment for the brain was reproducible in several experiments. Results from representative experiments performed with two different mixtures of libraries are shown in Fig. 1.

Sequencing of the inserts from 48 brain-localizing phage from library pool I revealed three dominant amino-acid sequence motifs. Peptides containing an SRL motif represented 54% of the clones, followed by a CENWWGDVC motif (29%). Other motifs that appeared more than once included CKDWGRIC, CVLRGGRC and CTRITESC. Many of these less common motifs shared the sequence RI/RL with the more common ones. Eight sequences were seen only once and probably represent background. From the library pool II phage, 25 sequences revealed only one motif, WRCVLREGPAGGCAWFNRHRL, which comprised 40% of the sequences.

The SRL tripeptide was found in several sequence contexts, indicating that the sequences were derived from a number of independent phage. Moreover, the DNA sequences of phage displaying the same peptide were in some cases not identical. The strong selection for predominant motifs and their internal diversity clearly shows that the peptide displayed by the phage, rather than some incidental mutant property of the phage, is responsible for the selective binding.

When tested as isolated phage, the CLSRLDAC, CNSRLHLRC, CENWWGDVC and WRCVLREGPAGGCAWFNRHRL phage each targeted the brain several-fold more effectively than the kidney. The brain/kidney ratios (number of phage recovered from brain divided by number of phage recovered from the same amount of kidney tissue) were about 8 for the CLSRLDAC and the CNSRLHLRC phage, 4 CENWWGDVC, and 9 for WRCVLREGPAGGCAWFNRHRL. A phage that had not been selected for brain or kidney binding gave a brain/kidney ratio of approximately 1.

Phage that would home selectively into the kidney were isolated from a mixture of the CX₃C and CX₃C libraries. The enrichment was three- to fivefold (data not shown). Phage carrying two motifs, CLPVASC and CGAREMC, constituted 60% of the 48 phage sequenced; CKGRSSAC appeared three times. The preferential binding to kidney relative to brain was highest, about sevenfold, with CLPVASC phage. Control phage again gave a kidney/brain ratio of about 1.

Immunohistochemical staining of the brain-binding phage displaying CLSRLDAC revealed staining within the brain capil-

laries (Fig. 2a). No preference for any part of the brain was seen. Injection of the kidney-binding CLPVASC phage did not cause staining of the brain capillaries (Fig. 2b). In contrast, the kidney-binding phage was found in the glomeruli and in between the tubules (Fig. 2c). Only slight staining was seen in the kidney with the brain-binding phage (Fig. 2d).

A soluble cyclic peptide was synthesized according to one of the brain-binding phage sequences, CLSSRLDAC. We chose to synthesize this peptide because it gave a slightly higher (8.2 versus 7.6) brain/kidney ratio than the most prevalent motif, CNSRLHLRC. The CLSSRLDAC peptide inhibited the preferential localization into the brain of the phage carrying the same

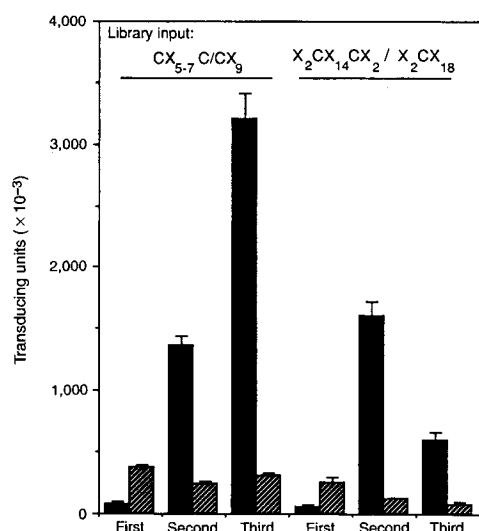


FIG. 1 Selective localization of phage to the brain. Two library pools, CX₅₋₇C/CX₉ (pool I)¹⁰⁻¹² and X₂CX₁₄CX₂, X₂CX₁₈ (pool II) (E. Koivunen and E.R., unpublished data) were injected in the tail vein. After 1–4 min the mice were killed and the phage rescued from tissues. Phage recovered from the brain were amplified and re-injected in two consecutive rounds. The number of phage (transducing units) recovered from brain (black bars) and kidney (hatched bars) tissue in a representative experiment with each library pool is shown. The number of phage recovered from different organs varied to some extent, but the ratios were consistent across experiments. The bars show standard error of the mean (s.e.m.) from plating in triplicate.

METHODS. The libraries were prepared as described¹⁰⁻¹³, and display mostly cyclic peptides, which often bind with higher affinity than non-cyclic peptides^{11,12}. Balb/c (2-month-old females; Jackson Laboratories, Bar Harbor, ME) were anaesthetized with Avertin (0.015 ml g⁻¹) and injected intravenously (tail vein) with a mixture of phage libraries containing 10¹⁶ (pool I) and 10¹⁴ (pool II) transducing units diluted in 200 µl DMEM. At the end of the experiment the mice were snap-frozen in liquid nitrogen, while in a state of deep anaesthesia. To recover the bound phage, the carcasses were partly thawed at room temperature, organs were removed, weighed, and ground in 1 ml DMEM-PI (DMEM containing the protease inhibitors phenyl methyl sulphonyl fluoride (1 mM), aprotinin (20 µg ml⁻¹) and leupeptin (1 µg ml⁻¹)). The tissue were washed three times with ice-cold DMEM-PI containing 1% BSA and incubated with 1 ml of bacteria for 1 h. NZY medium (10 ml) containing 0.2 µg ml⁻¹ tetracycline was added, the mixture was incubated in a 37 °C shaker for 1 h, and 200-µl portions were plated in agar plates in the presence of 40 µg ml⁻¹ tetracycline. About 200 individual colonies were grown separately for 16 h in 5 ml NZY medium containing 40 µg ml⁻¹ tetracycline. The bacterial cultures were then pooled and the amplified phage were injected into mice as described above.

FIG. 2 Immunohistochemical staining of phage in brain and kidney tissue. Phage were amplified individually and injected into mice. Tissue sections were prepared after perfusion of the mice through the heart with DMEM, and the organs were fixed in Bouin solution. An antibody against M13 (Pharmacia Biotech, Piscataway, NJ) was used for the staining, followed by a peroxidase-conjugated secondary antibody (Sigma, St Louis, MO). Brain-selective phage displaying CLSSRLDAC in brain (a) and in kidney (d), and the kidney-selective phage displaying CLPVASC in kidney (c) and in brain (b) are shown. Magnification: a, ×400; b–d, ×200.

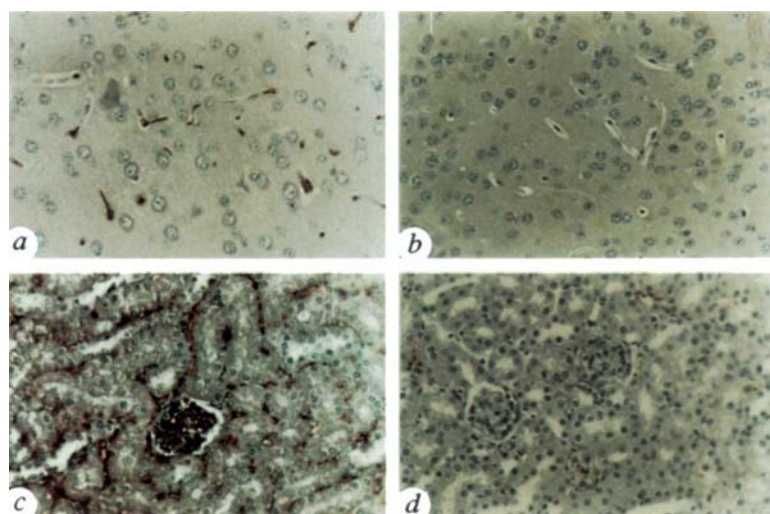
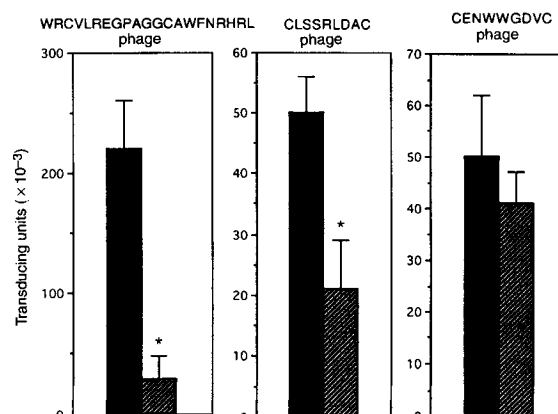


FIG. 3 Effect of the CLSSRLCAC synthetic peptide on the *in vivo* localization of phage. Brain-binding phage displaying CLSSRLDAC, CENWWGDVC and WRCVLREGPAGGCAWFNRHRL motifs were titrated to the same concentration and 10⁸ transducing units were injected into mice either on their own (black bars) or together with 500 µg of the CLSSRLDAC synthetic peptide (hatched bars). Shown is the number of phage (transducing units) recovered from the brain. The CLSSRLDAC peptide was synthesized and purified by high-performance liquid chromatography (Immunodynamics, La Jolla, CA). Bars show s.e.m. from triplicates. Asterisks indicate statistically significant differences (unpaired Student *t*-test, *P* < 0.05).



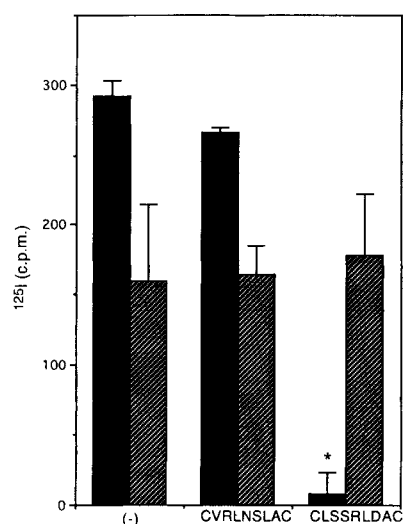


FIG. 4 Tissue localization of red blood cells coated with CLSSRLDAC peptide. Inhibition of brain localization by the corresponding soluble peptide. Iodinated CLSSRLDAC peptide was coupled to red blood cells and injected into the tail vein of mice in the presence or absence of unlabelled peptides. Radioactivity in perfused brain (black bars) and kidney (hatched bars) tissues is shown.

METHODS. The CLSSRLDAC peptide (1 mg) was labelled using the Bolton Hunter reagent (Amersham Life Science, Arlington Heights, IL), and purified by reversed-phase chromatography on Sep-Pak cartridges (Waters, Millipore, Milford, MA). The labelled peptide (100 µg) was coupled to 1 ml glutaraldehyde-stabilized sheep red blood cells (Sigma, St Louis, MO) according to the manufacturer's instructions. The coated cells (50 µl, 200,000 c.p.m.) were injected into the tail vein of mice in the presence or absence of 10 mM of unlabelled CLSSRLDAC peptide. An unrelated peptide, CVRLNSLAC, was used as a control. The mice were killed 2 min later, perfused through the heart with 50 ml DMEM, and their brain and kidneys were removed and assayed for radioactivity. The animals were treated in accordance with the Institute's Animal Facility Guidelines. Bars show s.e.m. from triplicates. Asterisks indicate statistically significant differences (unpaired Student's *t*-test, *P* < 0.05).

sequence and of the WRCVLREGPAGGCAWFNRHRL phage, but had no effect on the brain localization of the CENWWDGVC phage (Fig. 3). Thus the first two peptides, which were obtained from different libraries in two independent experiments, seem to bind to the same target molecule, possibly because of the similarity of the C terminus of the long peptide with the SRL motif. Differences in binding parameters may explain the greater susceptibility of the long motif to inhibition by the CLSSRLDAC peptide. The third peptide is likely to have a different target.

We also showed that CLSSRLDAC could target a particle other than the phage to the brain. Coupling the peptide onto the surface of red blood cells resulted in their accumulation in the brain to a greater extent than in the kidney (Fig. 4). Moreover, the brain localization of the red blood cells was blocked by co-injection of the soluble peptide, whereas the accumulation in the kidney was not affected (Fig. 4).

Future studies will be needed to identify the molecules to which the peptides bind in the brain and kidney. The sequences of the binding motifs are not helpful in this regard, because they do not reveal any significant similarities with known receptor ligands. Our initial attempts to identify the target molecule for the CLSSRLDAC peptide by affinity chromatography of brain extracts have not been successful, possibly because endothelial cell molecules would only be present as minor components in a brain extract. The receptors for the peptides are likely to be endothelial cell molecules, because the phage were allowed to circulate only for a few minutes, making it unlikely that the phage would have left the circulation. Moreover, immunohistochemical

staining of phage after injection showed that they remain in the lumen of blood vessels in the targeted organs.

To our knowledge this is the first time an *in vivo* selection procedure has been applied to a random library. So far we have targeted only two organs, the brain and the kidney, and were in each case able to recover organ-selective phage. This initial success suggests that it will be possible to apply this procedure to the identification of selective binding sequences for other organs as well, although organs that capture a large number of phage, such as liver and lung, may prove rather more troublesome. The method should be applicable to phage display libraries expressing larger proteins including the antibody variable binding region and the binding domains of specific ligands, as well as random libraries based on principles other than phage display; the only requirement is the ability to identify the compound in the tissue after the *in vivo* binding.

Organ-selective targeting molecules isolated from random libraries following the procedures described here may have a variety of uses. It may be possible to graft motifs to surface molecules of viruses or cells used in gene therapy. Other possibilities include their use in the preparation of drug conjugates or liposomes with specific targeting properties. Tumour vasculature, which undergoes active angiogenesis and contains specific markers^{14,15}, would be a particularly attractive future target, as it might allow therapies to be directed into tumours while sparing other tissues.

Received 28 December 1995; accepted 20 February 1996.

1. Fidler, I. J. & Hart, I. R. *Science* **217**, 998–1003 (1982).
2. Johnson, R. C. et al. *Cancer Res.* **51**, 394–399 (1991).
3. Springer, T. A. *Cell* **76**, 301–314 (1994).
4. Salmi, M. et al. *Proc. natn. Acad. Sci. U.S.A.* **89**, 11436–11440 (1992).
5. Bevilacqua, M. P., Stengelin, S., Gimbrone, M. A. & Seed, B. *Science* **243**, 1160–1165 (1989).
6. Siegelman, M. H., Rijn, M. & Weissman, I. L. *Science* **243**, 1165–1171 (1989).
7. Cepek, K. L. et al. *Nature* **372**, 190–193 (1994).
8. Rosen, S. D. & Bertozzi, C. R. *Curr. Opin. Cell Biol.* **6**, 663–673 (1994).
9. Johnson, R. C., Zhu, D., Augustin-Voss, H. G. & Pauli, B. U. *J. Cell Biol.* **121**, 1423–1432 (1993).
10. Smith, G. P. & Scott, J. K. *Meth. Enzym.* **217**, 228–257 (1993).
11. Koivunen, E., Wang, B. & Ruoslahti, E. *J. Cell Biol.* **124**, 373–380 (1994).
12. Koivunen, E., Wang, B., Dickinson, C. D. & Ruoslahti, E. *Meth. Enzym.* **245**, 346–369 (1994).
13. Pasqualini, R., Koivunen, E. & Ruoslahti, E. *J. Cell Biol.* **130**, 1189–1196 (1995).
14. Brooks, P. C. et al. *Cell* **79**, 1157–1164 (1994).
15. Rak, J. W., St. Croix, B. D. & Kerbel, R. S. *Anticancer Drugs* **6**, 3–18 (1995).

ACKNOWLEDGEMENTS. We thank E. Engvall, B. Öbrink, J. Smith, K. Vuori, W. Arap, W. Cavennee, E. Franco, E. Koivunen, H. Xu, S. Krajewski and R. Jenkins for advice. This work was supported by grants from the National Cancer Institute (E.R.). R.P. is the recipient of a fellowship from the Arthritis Foundation.

CORRESPONDENCE AND MATERIALS. Requests to be addressed to E.R. (e-mail address ruoslahti@lcrf.edu).

CORRECTION

Behavioural and cardiovascular effects of disrupting the angiotensin II type-2 receptor gene in mice

Lutz Hein, Gregory S. Barsh, Richard E. Pratt, Victor J. Dzau & Brian K. Kobilka

Nature **377**, 744–747 (1995)

REFERENCES 22 and 25 of this paper should be replaced as follows:

22. Nakajima, M. et al. *Proc. natn. Acad. Sci. U.S.A.* **92**, 10663–10667 (1995).
25. Viswanathan, M. & Saavedra, J. M. *Peptides* **13**, 783–786 (1992).